

日本国特許庁
JAPAN PATENT OFFICE

M. Kawato

3/19/04

Q 80487

10f1

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年 3月31日
Date of Application:

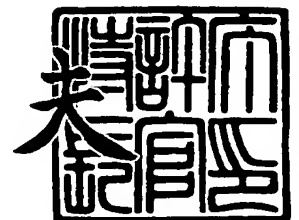
出願番号 特願2003-093459
Application Number:
[ST. 10/C]: [JP 2003-093459]

出願人 日本電気株式会社
Applicant(s):

2003年12月11日

特許庁長官
Commissioner,
Japan Patent Office

今井 康



出証番号 出証特2003-3102709

【書類名】 特許願

【整理番号】 35001195

【提出日】 平成15年 3月31日

【あて先】 特許庁長官 殿

【国際特許分類】 H04L 12/54
H04L 12/58

【発明者】

【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内

【氏名】 川戸 正裕

【特許出願人】

【識別番号】 000004237

【氏名又は名称】 日本電気株式会社

【代理人】

【識別番号】 100080816

【弁理士】

【氏名又は名称】 加藤 朝道

【電話番号】 045-476-1131

【手数料の表示】

【予納台帳番号】 030362

【納付金額】 21,000円

【その他】 国等の委託研究の成果に係る特許出願（新エネルギー・産業技術総合開発機構からの委託研究「情報通信基盤高度化プログラム（ネットワークコンピューティング技術の開発）」、産業活力再生特別措置法第 3 0 条の適用を受けるもの）

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9304371

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 分散型資源管理システムおよび分散型資源管理方法並びにプログラム

【特許請求の範囲】

【請求項 1】

1 つ又は複数の資源を管理する資源管理装置であって、

管理対象の資源を 1 つ又は複数指定してなる仮予約要求を受理した際に、前記管理対象の資源のうち空き状態である時間帯を仮予約された状態に変更して、有効期限を設けるとともに、前記仮予約要求の発行者を識別する情報を記憶部に格納する手段と、

前記管理対象の資源のうち、仮予約された時間帯が存在する資源について設けられる、待ち行列に、仮予約要求を格納する手段と、

前記仮予約された時間帯と、前記仮予約された時間帯に対応する仮予約識別情報（「仮予約 I D」という）を、前記仮予約要求に対する応答とする手段と、

を備えている、ことを特徴とする資源管理装置。

【請求項 2】

前記仮予約 I D を指定した仮予約取消要求を受理した場合、前記仮予約 I D に対応する仮予約を取り消して空き状態とする手段を備えている、ことを特徴とする請求項 1 記載の資源管理装置。

【請求項 3】

仮予約 I D と時間帯を指定したジョブ予約要求を受理した際に、前記ジョブ予約要求で指定される時間帯が、前記仮予約 I D に対応する仮予約された時間帯に含まれており、かつ、前記ジョブ予約要求の発行者と、前記仮予約要求の発行者とが互いに一致している場合に、前記ジョブ予約要求で指定される時間帯を、ジョブ予約された状態に、変更する手段を備えている、ことを特徴とする請求項 1 記載の資源管理装置。

【請求項 4】

前記仮予約された時間帯のうち、有効期限に達しているものを空き状態に戻す手段と、

前記待ち行列から仮予約要求を取り出し、空き状態になった時間帯を仮予約された状態に変更し、仮予約要求の発行者を識別する情報を格納する手段と、

前記待ち行列から取り出された仮予約要求の発行者に対して、前記仮予約された時間帯と、前記仮予約された時間帯に対応する仮予約IDを含む通知を送る手段と、

を備えている、ことを特徴とする請求項1記載の資源管理装置。

【請求項5】

1つ又は複数の資源を管理する資源管理装置と通信接続され、ジョブのスケジューリングを行うジョブスケジューラにおいて、

前記ジョブが利用する可能性のある資源のそれぞれについて、該資源を管理する前記資源管理装置に対して、仮予約要求を送る第1の手段と、

仮予約された資源を対象として、各ジョブを割り当てる資源と、実行する時間帯とを決める第2の手段と、

前記第2の手段によって決定された資源のそれぞれについて、該資源を管理する前記資源管理装置に対して、ジョブ予約要求を送る第3の手段と、

前記第1の手段によって仮予約された資源のそれぞれについて、該資源を管理する前記資源管理装置に対して、仮予約取消要求を送る第4の手段と、

を備えている、ことを特徴とするジョブスケジューラ。

【請求項6】

前記第1の手段で待ち行列に格納された仮予約要求がある場合に、予め定められた定時間が経過するか、前記待ち行列から取り出されたことを示す通知が到着してから、前記第2の手段に進む制御を行う手段を備えている、ことを特徴とする請求項5記載のジョブスケジューラ。

【請求項7】

請求項1乃至4のいずれかに記載の資源管理装置と、

請求項5又は6記載のジョブスケジューラと、

を備えた分散資源管理システム。

【請求項8】

1つ又は複数の資源の管理を行う少なくとも1つの資源管理装置と、

少なくとも 1 つの前記資源管理装置と通信接続される少なくとも 1 つのジョブスケジュール装置と、

を含み、

前記ジョブスケジュール装置は、前記ジョブで用いる資源の仮予約を行う手段を備え、

前記資源管理装置は、前記ジョブスケジュール装置により発行される、仮予約とジョブ予約との 2 種の予約要求をそれぞれ受理する手段を備え、

前記仮予約は、資源の予約状況の取得と資源のロックを組み合わせた操作であって、前記資源管理装置の管理下にある資源のいずれかを指定して実行され、

前記資源管理装置は、前記ジョブスケジュール装置により受け取った前記仮予約要求の対象の資源に対して、他の仮予約が存在せず、かつ、空き時間が存在する場合には、当該時間帯を仮予約された状態に設定し、前記仮予約された時間帯を、前記仮予約要求の応答とする手段と、

前記仮予約された時間帯に、有効期限と、該当する仮予約要求を発行したジョブスケジュール装置の情報とを記録する手段と、

を有し、

前記ジョブ予約は、指定されたジョブの実行のために資源を事前予約する操作であって、前記資源管理装置が管理する資源のうちの 1 つと、前記資源における時間帯を指定したジョブ予約要求として実行され、

前記資源管理装置は、前記ジョブ予約要求で指定された時間帯がすでに仮予約された時間帯の一部であり、かつ、前記ジョブ予約要求の発行者と仮予約要求の発行者とが互いに一致する場合に、ジョブ予約要求を受理し、指定された時間帯をジョブ予約された状態とする手段と、

ジョブ予約で指定された開始時刻に達すると、当該資源をジョブの実行のために割り当てる制御、及び／又は、前記ジョブ予約で指定された終了時刻になると当該の資源のジョブへの割り当てを停止する制御、を行う手段と、

を含む、ことを特徴とする分散型資源管理システム。

【請求項 9】

前記資源管理装置は、前記予約要求に対応して、前記ジョブスケジュール装置

により発行される、仮予約取消要求とジョブ取消要求の2種の取消要求を受理する、ことを特徴とする請求項8記載の分散型資源管理システム。

【請求項10】

前記ジョブスケジュール装置に通信接続する少なくとも1つのユーザ端末をさらに備え、

前記ユーザ端末から、前記ジョブスケジュール装置にジョブが投入され、

前記ジョブスケジュール装置は、

前記ユーザ端末から前記投入されたジョブを各資源に割り当てる手段を備え、

前記ジョブスケジュール装置による、前記資源の割り当て処理が、

前記資源管理装置に対する、前記仮予約要求の送信処理と、

前記仮予約に成功した資源を対象とするジョブの実行スケジュールの作成処理と、

前記実行スケジュールに従ったジョブ予約要求を前記資源管理装置に対して送信する処理と、

の少なくとも1つを含む、ことを特徴とする請求項8記載の分散型資源管理システム。

【請求項11】

前記資源管理装置は、前記資源管理装置の管理対象とされる資源のそれぞれについて、どの時間帯にどのジョブが予約されているかの情報を記録するスケジュール記憶部と、

前記資源管理装置の管理対象とされる資源のそれぞれについて、待ち行列が設けられ、前記待ち行列には、前記資源を対象とする待機中の仮予約要求が格納される仮予約要求記憶部と、

前記ジョブスケジュール装置からの仮予約要求を受信し、前記スケジュール記憶部又は、前記仮予約要求記憶部に、仮予約についての情報を記録する仮予約要求受信手段と、

前記ジョブスケジュール装置からのジョブ予約要求を受信し、前記スケジュール記憶部に、ジョブ予約の情報を記録するジョブ予約要求受信手段と、

前記スケジュール記録部に記録された予約ジョブ及び仮予約の情報を監視し、

前記予約ジョブ及び前記仮予約の情報に従った、資源の割り当て・解放、及び／又は、前記仮予約の破棄の処理を行うスケジュール管理手段と、

を備えている、ことを特徴とする請求項 8 記載の分散型資源管理システム。

【請求項 12】

前記資源管理装置は、前記ジョブスケジュール装置からの仮予約取消要求を受信し、前記スケジュール記録部又は前記仮予約要求記憶部から該当する仮予約の情報を破棄する仮予約取消要求受信手段と、

前記ジョブスケジュール装置からのジョブ予約取消要求を受信し、前記スケジュール記録部から該当するジョブ予約の情報を破棄する処理と、ジョブに割り当てられた資源を解放する処理を行うジョブ取消要求受信手段と、

を備えている、ことを特徴とする請求項 11 記載の分散型資源管理システム。

【請求項 13】

前記ジョブスケジュール装置は、前記ユーザ端末により投入されたジョブの情報が一時的に格納される待機ジョブ記憶部と、

各資源ごとに、予約が確定したジョブの情報が格納されるスケジュール記憶部と、

前記ユーザ端末からのジョブ実行要求を受信し、前記ジョブ実行要求の内容を待機ジョブ情報記憶部に記録するジョブ実行要求受信手段と、

前記待機ジョブ記憶部に記録されたジョブを、定期的に取り出し、前記ジョブを資源に割り当てる手段であって、前記ジョブの資源への割り当てには、前記資源管理装置への仮予約要求の発行、ジョブ実行スケジュールの作成、および、前記資源管理装置へのジョブ予約要求の発行が含まれる、ジョブスケジューリング手段と、

を備えている、ことを特徴とする請求項 8 記載の分散型資源管理システム。

【請求項 14】

前記ジョブスケジュール装置は、前記ユーザ端末からのジョブ取消要求を受信し、前記待機ジョブ記憶部に記憶された待機ジョブまたは前記スケジュール記憶部に登録された該当するジョブを破棄するとともに、前記資源管理装置に対するジョブ予約取消要求の送信を行うジョブ取消要求受信手段を備えている、ことを

特徴とする請求項 13 記載の分散型資源管理システム。

【請求項 15】

前記資源管理装置は、前記ジョブスケジュール装置から受信したジョブ予約要求に従って、管理下の資源を、前記ジョブの実行のために割り当てる手段と、

前記資源管理装置は、すでに仮予約された資源を対象とする仮予約要求を受け取ると、前記仮予約要求を資源ごとに設けられる待ち行列に格納する手段と、
を備え、

仮予約取消要求または有効期限切れのために仮予約が破棄されると、破棄された仮予約と同一の資源を対象とする仮予約要求が待ち行列から取り出され、有効な仮予約となり、

前記資源管理装置は、前記待ち行列から仮予約要求が取り出されると、前記仮予約要求を発行した前記ジョブスケジュール装置に通知を送る手段を備え、

前記ジョブスケジュール装置は、前記ジョブスケジュール装置が発行した仮予約要求のいずれかが前記待ち行列に入ると、前記仮予約が有効となるのを待ってから、実行スケジュールの作成を開始するように制御する手段を備えている、ことを特徴とする請求項 8 記載の分散型資源管理システム。

【請求項 16】

前記ジョブスケジュール装置は、予め定められた定時間内に、仮予約要求が有効にならない場合には、その時点で、前記待ち行列に残っている仮予約を無視して、実行スケジュールの作成に進むように制御する手段を備えている、ことを特徴とする請求項 14 記載の分散型資源管理システム。

【請求項 17】

1 つ又は複数の資源を管理する資源管理装置による資源管理の方法であって、
前記資源管理装置が、管理対象の資源を 1 つ又は複数指定してなる仮予約要求を受理した際に、前記管理対象の資源のうち空き状態である時間帯を仮予約された状態に変更して、有効期限を設けるとともに、前記仮予約要求の発行者を識別する情報を記憶部に格納する工程と、

前記資源管理装置が、前記管理対象の資源のうち、仮予約された時間帯が存在する資源についての待ち行列に、仮予約要求を格納する工程と、

前記資源管理装置が、前記仮予約された時間帯と、前記仮予約された時間帯に対応する仮予約識別情報（「仮予約ID」という）を、前記仮予約要求に対する応答とする工程と、

を含む、ことを特徴とする資源管理方法。

【請求項18】

前記資源管理装置が、前記仮予約IDを指定した仮予約取消要求を受理した場合に、前記仮予約IDに対応する仮予約を取り消して空き状態とする工程を含む、ことを特徴とする請求項17記載の資源管理方法。

【請求項19】

前記資源管理装置が、前記仮予約IDと時間帯を指定したジョブ予約要求を受理した際に、前記ジョブ予約要求で指定される前記時間帯が、前記仮予約IDに対応する仮予約された時間帯に含まれ、かつ、前記ジョブ予約要求の発行者と、前記仮予約要求の発行者とが互いに一致している場合に、前記ジョブ予約要求で指定される時間帯をジョブ予約された状態に変更する工程を含む、ことを特徴とする請求項17記載の資源管理方法。

【請求項20】

前記資源管理装置が、仮予約された時間帯のうち、有効期限に達しているものを空き状態に戻す工程と、

前記資源管理装置が、前記待ち行列から仮予約要求を取り出し、空き状態になった時間帯を仮予約された状態に変更し、仮予約要求の発行者を識別する情報を格納する工程と、

前記資源管理装置が、前記待ち行列から取り出された仮予約要求の発行者に対して、仮予約された時間帯とそれに対応する仮予約IDを含む通知を送る工程と、

を含む、ことを特徴とする請求項17記載の資源管理方法。

【請求項21】

1つ又は複数の資源を管理する資源管理装置と通信接続され、ジョブのスケジューリングを行うジョブスケジューラが、前記ジョブが利用する可能性のある資源のそれぞれについて、該資源を管理する前記資源管理装置に対して、仮予約要

求を送る工程と、

前記ジョブスケジューラが、仮予約された資源を対象として、各ジョブを割り当てる資源と実行する時間帯とを決める工程と、

前記ジョブスケジューラによって決定された資源のそれぞれについて、該資源を管理する前記資源管理装置に対して、ジョブ予約要求を送る工程と、

前記仮予約された資源のそれぞれについて、該資源を管理する前記資源管理装置に対して、仮予約取消要求を送る工程と、

を含み、

前記各工程により、ジョブの資源への割り当てを決める、ことを特徴とする請求項 17 記載の資源管理方法。

【請求項 22】

前記ジョブスケジューラは、前記仮予約要求を送る工程において、前記待ち行列に格納された仮予約要求がある場合に、予め定められた定時間が経過するか、前記待ち行列から取り出されたことを示す通知が到着してから、前記資源と実行する時間帯とを決める工程に進む制御が行われる、ことを特徴とする請求項 21 記載の資源管理方法。

【請求項 23】

1 つ又は複数の資源の管理を行う、少なくとも 1 つの資源管理装置と、

少なくとも 1 つの前記資源管理装置と通信接続される、少なくとも 1 つのジョブスケジューラ装置と、を含む分散システムの資源管理方法であって、

前記ジョブスケジューラ装置が、前記ジョブで用いる資源の仮予約を行う工程を含み、

前記資源管理装置が、前記ジョブスケジューラ装置により発行される、仮予約とジョブ予約との 2 種の予約要求を受理する工程を含み、

前記仮予約は、資源の予約状況の取得と、資源のロックを組み合わせた操作であって、前記資源管理装置の管理下にある資源のいずれかを指定して実行され、

前記資源管理装置が、前記ジョブスケジューラ装置より受け取った前記仮予約要求の対象の資源に対して、他の仮予約が存在せず、かつ、空き時間が存在する場合には、当該時間帯を仮予約された状態に設定し、前記仮予約された時間帯を

、前記仮予約要求に対する応答とする工程と、

前記資源管理装置が、前記仮予約された時間帯に、有効期限と、該当する仮予約要求を発行したジョブスケジュール装置の情報とを記録する工程と、

を有し、

前記ジョブ予約は、指定されたジョブの実行のために資源を事前予約する操作であって、前記資源管理装置が管理する資源のうちの1つと、前記資源における時間帯を指定したジョブ予約要求として実行され、

前記資源管理装置が、前記ジョブ予約要求で指定された時間帯がすでに仮予約された時間帯の一部であり、かつ、前記ジョブ予約要求の発行者と前記仮予約要求の発行者とが互いに一致する場合に、前記ジョブ予約要求を受理し、指定された時間帯をジョブ予約された状態とする工程と、

前記資源管理装置が、ジョブ予約で指定された開始時刻に達すると、当該資源をジョブの実行のために割り当てる制御を行い、及び／又は、前記ジョブ予約で指定された終了時刻になると、当該の資源のジョブへの割り当てを停止する制御を行う工程と、

を含む、ことを特徴とする分散型資源管理方法。

【請求項 24】

前記資源管理装置は、前記予約要求に対応して、前記ジョブスケジュール装置により発行される、仮予約取消要求とジョブ取消要求の2種の取消要求を受理する工程を含む、ことを特徴とする請求項 23 記載の分散型資源管理方法。

【請求項 25】

前記ジョブスケジュール装置に接続するユーザ端末から、前記ジョブスケジュール装置にジョブが投入される工程と、

前記ジョブスケジュール装置は、前記ユーザ端末から、前記投入されたジョブを、各資源に割り当てる処理を行う工程と、

を有し、

前記ジョブスケジュール装置による前記資源の割り当て処理は、

前記資源管理装置に対する、仮予約要求の送信処理と、

前記仮予約に成功した資源を対象とするジョブの実行スケジュールの作成処理

と、

前記実行スケジュールに従ったジョブ予約要求の前記資源管理装置への送信処理と、

を含む、ことを特徴とする請求項 23 記載の分散型資源管理方法。

【請求項 26】

前記資源管理装置は、前記ジョブスケジュール装置からの仮予約要求を受信した際に、

管理下の資源のそれぞれについてどの時間帯にどのジョブが予約されているかの情報を記録するスケジュール記憶部、又は、管理下の資源のそれぞれについて設けられる待ち行列に前記資源を対象とする待機中の仮予約要求が格納される仮予約要求記憶部に、仮予約についての情報を記録する工程と、

前記資源管理装置が、前記ジョブスケジュール装置からのジョブ予約要求を受信し、前記スケジュール記憶部に、ジョブ予約の情報を記録する工程と、

前記資源管理装置が、前記スケジュール記録部に記録された予約ジョブ及び仮予約の情報を監視し、前記予約ジョブ及び仮予約の情報に従った資源の割り当てと解放、及び／又は、仮予約の破棄の処理を行う工程と、

を含む、ことを特徴とする請求項 23 記載の分散型資源管理方法。

【請求項 27】

前記資源管理装置が、前記ジョブスケジュール装置からの仮予約取消要求を受信し、前記スケジュール記録部又は前記仮予約要求記憶部から、該当する仮予約の情報を破棄する工程と、

前記資源管理装置が、前記ジョブスケジュール装置からのジョブ予約取消要求を受信し、前記スケジュール記録部から該当するジョブ予約の情報を破棄し、及び、ジョブに割り当てられた資源を解放する処理を行う工程と、

を含む、ことを特徴とする請求項 23 記載の分散型資源管理方法。

【請求項 28】

前記ジョブスケジュール装置は、前記ジョブスケジュール装置に接続するユーザ端末からのジョブ実行要求を受信し、前記ジョブ実行要求の内容を、前記ユーザ端末により投入されたジョブの情報が一時的に格納される待機ジョブ情報記憶

部に記録する工程と、

前記ジョブスケジュール装置が、前記待機ジョブ記憶部に記録されたジョブを取り出して、前記ジョブを資源に割り当てる工程と、

を有し、

前記ジョブの資源への割り当てには、前記資源管理装置への仮予約要求の発行、ジョブ実行スケジュールの作成、および、前記資源管理装置へのジョブ予約要求の発行が含まれる、ことを特徴とする請求項 2 3 記載の分散型資源管理方法。

【請求項 2 9】

前記ジョブスケジュール装置は、前記ジョブスケジュール装置に接続するユーザ端末からジョブ取消要求を受信した際に、前記待機ジョブ記憶部に記憶された待機ジョブ、または、各資源ごとに予約が確定したジョブの情報が格納されるスケジュール記憶部に登録された該当するジョブを破棄するとともに、前記資源管理装置に対してジョブ予約取消要求を送信する工程を含む、ことを特徴とする請求項 2 8 記載の分散型資源管理方法。

【請求項 3 0】

前記資源管理装置は、前記ジョブスケジュール装置から受信したジョブ予約要求に従って、管理下の資源を、前記ジョブの実行のために割り当てる工程と、

前記資源管理装置は、すでに仮予約された資源を対象とする仮予約要求を受け取ると、前記仮予約要求を資源ごとに設けられる待ち行列に格納する工程と、

を有し、

仮予約取消要求または有効期限切れのために仮予約が破棄されると、破棄された仮予約と同一の資源を対象とする仮予約要求が待ち行列から取り出され、有効な仮予約となり、

前記資源管理装置が、前記待ち行列から仮予約要求が取り出されると、前記仮予約要求を発行した前記ジョブスケジュール装置に通知を送る工程を有し、

前記ジョブスケジュール装置は、前記ジョブスケジュール装置が発行した仮予約要求のいずれかが前記待ち行列に入ると、前記仮予約が有効になるのを待ってから、実行スケジュールの作成を開始する工程を有する、ことを特徴とする請求項 2 3 記載の分散型資源管理方法。

【請求項 3 1】

前記ジョブスケジュール装置は、予め定められた定時間内に前記仮予約要求が有効にならない場合には、その時点で前記待ち行列に残っている仮予約を無視して実行スケジュールの作成に進む工程を含む、ことを特徴とする請求項 2 3 記載の分散型資源管理方法。

【請求項 3 2】

1 つ又は複数の資源を管理する少なくとも 1 つの資源管理装置と、
前記資源管理装置に通信接続され、前記資源を利用するジョブのスケジュールを行う少なくとも 1 つのジョブスケジュール装置と、
前記ジョブスケジュール装置に通信接続する少なくとも 1 つのユーザ端末と、
を含むシステムであって、
資源の利用を予約するために、前記資源管理装置に対して仮予約とジョブ予約の 2 段階を設け、
前記資源管理装置は、前記ジョブスケジュール装置からの仮予約要求を受け取り対応する資源を、仮予約された状態に設定する手段と、
仮予約されている資源に対して、仮予約要求が発行された場合、仮予約を待ち行列に格納する手段と、
を備え、
前記ユーザ端末は、前記ジョブスケジュール装置に対して、ジョブの実行を要求する手段を備え、
前記ジョブスケジュール装置は、前記資源管理装置に対して仮予約を発行した後に、ジョブを割り当てる資源を決定し、ジョブ予約を、前記資源管理装置に発行することで、ジョブの実行に必要な資源を、予約する手段を備えている、
ことを特徴とする分散型資源管理システム。

【請求項 3 3】

1 つ又は複数の資源を管理する少なくとも 1 つの資源管理装置と、
前記資源管理装置に通信接続され、前記資源を利用するジョブのスケジュールを行う、少なくとも 1 つのジョブスケジュール装置と、
前記ジョブスケジュール装置に通信接続する少なくとも 1 つのユーザ端末と、

を有する分散システムの資源管理方法であって、

資源の利用を予約するために、前記資源管理装置に対して仮予約とジョブ予約の2段階を設け、

前記ジョブスケジュール装置から前記資源管理装置に対して仮予約を行った後に、ジョブ予約が行われる工程と、

前記仮予約された資源に対して、前記ジョブスケジュール装置から、仮予約要求が発行された場合、前記資源管理装置では、前記仮予約要求を待ち行列に入れる工程と、

を有し、

前記ユーザ端末が、前記ジョブスケジュール装置に対して、ジョブの実行を要求する工程を有し、

前記ジョブスケジュール装置は、前記資源管理装置に対して仮予約を発行した後に、ジョブを割り当てる資源を決定し、前記資源管理装置に対してジョブ予約を発行することで、ジョブの実行に必要な資源を予約する、ことを特徴とする分散型資源管理方法。

【請求項 34】

1つ又は複数の資源を管理する資源管理装置を構成するコンピュータに、

管理対象の資源を1つ又は複数指定してなる仮予約要求を受理した際に、前記管理対象の資源のうち、空き状態である時間帯を、仮予約された状態に変更して、有効期限を設けるとともに、前記仮予約要求の発行者を識別する情報を記憶部に格納する処理と、

前記管理対象の資源のうち、仮予約された時間帯が存在する資源についての待ち行列に、仮予約要求を格納する処理と、

前記仮予約された時間帯と、前記仮予約された時間帯に対応する仮予約識別情報（「仮予約ID」という）を応答する処理と、

を実行させるプログラム。

【請求項 35】

請求項 34 記載のプログラムであって、

前記仮予約IDを指定した仮予約取消要求を受理した場合、前記仮予約IDに

対応する仮予約を取り消して空き状態にする処理を、前記コンピュータに実行させるプログラム。

【請求項 36】

請求項 34 記載のプログラムであって、

仮予約 ID と時間帯を指定したジョブ予約要求を受理した際に、前記ジョブ予約要求で指定される時間帯が、前記仮予約 ID に対応する仮予約された時間帯に含まれ、かつ、前記ジョブ予約要求の発行者と、前記仮予約要求の発行者とが互いに一致している場合に、前記ジョブ予約要求で指定される時間帯を、ジョブ予約された状態に変更する処理を、前記コンピュータに実行させるプログラム。

【請求項 37】

請求項 34 記載のプログラムであって、

前記仮予約された時間帯のうち、有効期限に達しているものを空き状態に戻す処理と、

前記待ち行列から仮予約要求を取り出し、空き状態になった時間帯を仮予約された状態に変更し、仮予約要求の発行者を識別する情報を格納する処理と、

待ち行列から取り出された仮予約要求の発行者に対して、前記仮予約された時間帯とそれに対応する仮予約 ID を含む通知を送る処理と、を前記コンピュータに実行させるプログラム。

【請求項 38】

請求項 34 記載の資源管理装置を構成する前記コンピュータと通信接続し、ジョブのスケジューリングを行うジョブスケジューラを構成するコンピュータに、

前記ジョブが利用する可能性のある資源のそれぞれについて、該資源を管理する前記資源管理装置に対して、仮予約要求を送る第 1 の処理と、

仮予約された資源を対象として、各ジョブを割り当てる資源と、実行する時間帯とを決める第 2 の処理と、

前記第 2 の処理によって決定された資源のそれぞれについて、該資源を管理する前記資源管理装置に対して、ジョブ予約要求を送る第 3 の処理と、

前記第 1 の処理によって仮予約された資源のそれぞれについて、該資源を管理する前記資源管理装置に対して、仮予約取消要求を送る第 4 の処理と、

を実行させるプログラム。

【請求項 39】

請求項 38 記載のプログラムにおいて、

前記第 1 の処理で前記待ち行列に格納された仮予約要求がある場合に、予め定められた定時間が経過するか、あるいは、前記待ち行列から取り出されたことを示す通知が到着してから、前記第 2 の処理に進むように制御する処理を、前記ジョブスケジューラを構成する前記コンピュータに実行させるプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は分散型資源管理システムおよび分散型管理方法に関し、特にジョブ単位での資源予約を行う分散型資源管理システムおよび分散型資源管理方法とコンピュータ・プログラムに関する。

【0002】

【従来の技術】

分散コンピューティングシステム（分散システム）は、コンピュータや記憶装置、ネットワーク等の多様な資源から構成される。このため、ある仕事（ジョブ）に必要なサービス品質（Quality of Service, QoS）を得るためには、複数の資源に渡ってサービス品質を保証するような資源管理方式が必要となる。特に、独立して管理される複数の部分システム（以下、「ドメイン」と呼ぶ）間で資源を共有するグリッドコンピューティングでは、システム全体の資源を一元的に管理することが許容されないことが多い。この種のシステムでは、複数の独立した資源管理機構を利用した QoS 保証機能が必要になる。

【0003】

分散システムにおいて QoS を保証する従来の技術として、資源の事前予約（以下、特に記載がない場合には予約は事前予約を指す）が知られている。事前予約とは、ジョブの実行に必要な個々の資源について、一定の時間帯において QoS を保証する操作である。

【0004】

図1を参照して、資源の事前予約の例を説明する。ここでは、記憶装置Bから入力したデータを計算機Aにより解析し、その結果を記憶装置Cに格納するという処理を仮定する（図1（A）参照）。

【0005】

一連の処理を滞りなく行うために、利用するそれぞれの資源、すなわち計算機A、記憶装置B、記憶装置Cに対して、事前予約を行う。その際、計算機Aにおけるデータ入力処理に関する予約と、記憶装置Bにおけるデータ読み出し処理に関する予約を同じ時間帯に対して行う必要がある（図1（B）参照）。計算機Aにおけるデータ出力処理と記憶装置Cにおけるデータ書き込み処理についても同様である。

【0006】

事前予約機能を持つ従来の資源管理システムの一例が知られている（例えば特許文献1参照）。図2に示すように、この従来の資源管理システムは、複数のユーザ端末400と、複数の資源300と、複数の接続制御装置700と、資源管理装置800とから構成されている。このような構成を有する従来の資源予約システムは、次のように動作する。

【0007】

（1）ユーザ端末400は、複数の接続制御装置700のどれかを介して資源管理装置800に対して予約要求を送る。

【0008】

（2）上記（1）の予約要求の結果、ユーザ端末400に予約証明書が渡される。

【0009】

（3）予約された時刻に達すると、ユーザ端末400は、予約された資源300に対して資源利用要求を送る。この資源利用要求には、（2）で発行された予約証明書が含まれる。

【0010】

上記特許文献1記載のシステムでは、資源の事前予約は、単一の資源管理装置800が一元的に管理する。このため、上記特許文献1に記載されているシステ

ムは、複数の資源管理装置によって資源が分散管理される分散システムに適用することはできない。

【0011】

資源が分散管理されることを前提とした別の資源管理技術の一例も知られている（例えば非特許文献1参照）。

【0012】

図3に示すように、上記非特許文献1記載の資源予約システムは、複数のユーザ端末400と、複数の資源300と、スケジューラ500（非特許文献1中では“co-reservation agent”と呼ばれている）と、資源情報サービス600とから構成されている。各資源300は、内部に資源管理部を備える。このような構成を有する従来の資源予約システムは次のように動作する。

【0013】

（1）ユーザ端末400は、複数のスケジューラ500のどれかに対して資源予約要求を送る。

【0014】

（2）スケジューラ500は、資源情報サービス600に問い合わせ、資源の予約状況を取得する。

【0015】

（3）スケジューラ500は、上記（2）で取得した予約状況に基づいて予約する資源を決定し、予約先の資源の資源管理部に対して資源予約要求を発行する。割り当てる資源が複数ある場合、それぞれの資源の資源管理部に対して同様に資源予約要求を発行する。

【0016】

（4）上記（3）の資源予約要求の結果、予約に対応する予約IDが発行され、上記（1）のユーザ端末400に渡される。

【0017】

（5）上記（1）のユーザ端末400は、上記（3）で予約されたそれぞれの資源の資源管理部に対して資源利用要求を発行する。この資源利用要求には、上記（4）で取得した予約IDが含まれる。

【0018】

なお、複数のクライアント装置から共有リソースを予約するアーキテクチャとして、開始時刻、期間、繰り返し時間シーケンスを含む予約エントリをリソース要求として、クライアント装置からAV/C掲示板に送信し、リソース要求は予約衝突を装置及びユーザに通知するためのリソースカレンダーとして組織化され、上記リソースをクライアント装置に割り当てるリソーススケジュールコントローラを備えたシステムが知られている（例えば特許文献2参照）。またリソースの仮予約を行うシステムとして、第1の端末装置は、リソース予約の要求メッセージを送信し、各ノード装置は、要求メッセージの受信時、予約可能か否か判定し、可能ならば仮予約を行って、次装置へ要求メッセージを送信し、不可の場合、前装置の予約否の応答メッセージを送信し、第2の端末装置は、要求メッセージ受信時に、データ通信に応じられるか否かを判定し、前装置に予約可又は予約否の応答メッセージを送信し、各ノード装置は予約可の応答メッセージ受信時に仮予約を本予約に変更し前装置に予約可の応答メッセージを送信するリソース予約方法も知られている（例えば特許文献3）。

【0019】

なお、後述されるジョブスケジューリングについて公知のアルゴリズムが参照される（例えば非特許文献2）。

【0020】

【特許文献1】

特開2000-259537号公報（第5頁、第1図）

【特許文献2】

特表2003-500961号公報（第14-17頁、第1図）

【特許文献3】

特開2002-185491号公報（第3-5頁、第1図）

【非特許文献1】

Ian Foster et.al., "A Distributed Resource Management Architecture that Supports Advance Reservations and Co-Allocation", International Workshop on Quality of Service 99., 1999

【非特許文献 2】

"Heuristic Algorithms for Scheduling Independent Tasks on Nonidentical Processors," The Journal of the ACM, Volume 24 Issue 2, 1977

【0021】

【発明が解決しようとする課題】

ところで、複数の資源管理装置が存在する分散システムにおいて、多数のジョブを一度に実行すると、資源の利用効率が低下する、という問題点を有している。その理由を以下に説明する。

【0022】

資源の事前予約を利用して複数のジョブを実行する方法として、

(i) ジョブが使う可能性のある資源を事前に予約した後、ジョブを割り当てる資源と時間帯を予約に成功した資源から選ぶ方法と、

(ii) ジョブを割り当てる資源と時間帯を決定した後、決定した個々の資源と時間帯について、事前予約する方法と、

の2種類があり得る。

【0023】

上記(i)の方法では、実際にはジョブを実行しない時間まで資源を予約することになり、上記(ii)に比べて資源の利用効率が低下する。さらに、予約した資源に対して利用要求を発行する予定のユーザ端末に障害が発生した場合、予約された資源が長時間誰にも使えない状態になる、という問題がある。

【0024】

一方、上記(ii)の方法では、予約しようとする資源がすでに他のユーザにより予約されていることがあり得る。このような場合に、予約の失敗が発生し、適切な資源の組み合わせが利用できない可能性がある。

【0025】

予約の失敗が発生するのは、ジョブの割り当てを決定するスケジューラ(図3の500)が最新の資源の予約状況を保持していない場合である。資源の予約状況は、資源情報サービス(図3の600)に問い合わせることで取得することができるが、資源情報サービスから得られる情報は、必ずしも最新の情報を反映し

ているとは限らない。仮に、個々の資源に直接予約状況を問い合わせるとしても、多数の資源に対して問い合わせるのに、長い時間を要し、その間に、状況が変化することがあり得る。また、ジョブの資源への割り当てを決めるジョブスケジューリング処理や、予約の実行の途中に予約状況が変化することもあり得る。

【0026】

予約の失敗が問題になる場合として、複数の資源を同時に使うジョブや、複数のジョブ間に依存関係がある場合が挙げられる。これらのジョブでは、一部の資源についてのみ予約に失敗すると、再度スケジューリングして残りの資源を予約する必要が生じる。この場合、最初に予約に成功した資源と、後で予約に成功した複数の資源の組み合わせが、当初可能であった組み合わせよりも、不適切な組み合わせしか利用できないことがあり得る。

【0027】

予約失敗による資源の利用効率の低下を、図4を用いて、説明する。この例では、広域ネットワーク（WAN）を通して2つの計算機クラスタA、Bが接続されている。各クラスタは、それぞれ8台ずつのノードからなる。このシステム上で、あるスケジューラが、4つのノードを使用する並列ジョブ1と、8つのノードを使用する並列ジョブ2についての資源を予約することを考える。

【0028】

前記スケジューラが把握している予約状況は、各クラスタとも予約は存在しないというものである。この状況では、スケジューラによるジョブの割り当ての候補として、クラスタAのうちの4ノードにジョブ1を、クラスタBの8ノードにジョブ2を割り当てるというものがあり得る（図4（A））。しかし、実際には、クラスタBのうちの4ノードに別のジョブ3がすでに予約されているとする（図4（B））。この場合、ジョブ1全体と、ジョブ2のうちのクラスタAに割り当てられる4ノードについてのみ予約が成功する。前記スケジューラは、予約に失敗したジョブ2の一部分について再度スケジューリングし、その結果クラスタAのうちの4ノードが選択されたとする（図4（C））。この結果、ジョブ2はクラスタAとクラスタBにまたがって割り当てられることになる。もし図4（B）の予約状況が事前に分かっていたら、ジョブ2をクラスタAにのみ割り当てること

が可能であり、WANを通した通信の必要はない。すなわち、予約の失敗によってネットワーク資源の利用効率が低下している。

【0029】

上記の例は、ネットワーク資源の利用効率が低下する場合であるが、他の種類の資源の利用効率が低下することもあり得る。例えば、均一な性能の計算機を割り当てることが望ましい並列ジョブに、不均一な性能の計算機しか割り当てられないことがあり得る。また、並列ジョブについてのみだけでなく、依存関係のある複数のジョブについても資源の利用効率の低下が起こりうる。例えば、あるジョブによる出力を別のジョブが入力とする場合、これらのジョブは同一の資源または同一のローカルエリアネットワーク（LAN）で接続された資源に割り当てられることが望ましいが、一方のジョブだけ予約に失敗した場合に、予約に失敗したジョブの割り当てが可能な資源が最初に予約した資源とは異なるLANにしかないことがあり得る。

【0030】

すなわち、複数の資源を同時に使うジョブや依存関係のあるジョブについて、資源の予約状況が正確に把握できないために、本来可能である資源の組み合わせよりも非効率な資源の組み合わせが割り当てられる。

【0031】

したがって、本発明の目的は、資源が複数の資源管理装置により分散管理される分散システムにおいて、正確な資源の予約状況に基づくジョブのスケジューリングを可能にし、各ジョブに適切な資源の組み合わせを実現する装置と方法並びにプログラムを提供することである。

【0032】

【課題を解決するための手段】

前記目的を達成する本発明の1つのアスペクトに係る装置は、1つ又は複数の資源を管理する資源管理装置であって、管理対象の資源を1つ又は複数指定してなる仮予約要求を受理した際に、前記管理対象の資源のうち、空き状態である時間帯を、仮予約された状態に変更し、有効期限を設けるとともに、前記仮予約要求の発行者を識別する情報を格納する手段と、前記管理対象の資源のうち、仮予

約された時間帯が存在する資源についての待ち行列に、仮予約要求を格納する手段と、前記仮予約された時間帯と、前記仮予約された時間帯に対応する仮予約 ID を前記仮予約要求の応答とする手段と、を備えている。

【0033】

本発明に係る装置において、前記仮予約 ID を指定した仮予約取消要求を受理した場合、前記仮予約 ID に対応する仮予約を取り消して空き状態とする構成としてもよい。

【0034】

本発明において、仮予約 ID と時間帯を指定したジョブ予約要求を受理した際に、指定される前記時間帯が、前記仮予約 ID に対応する仮予約された時間帯に含まれ、かつ、ジョブ予約要求の発行者と、仮予約要求の発行者が一致している場合に、指定される時間帯をジョブ予約された状態に変更する構成としてもよい。

【0035】

本発明において、仮予約された時間帯のうち、有効期限に達しているものを空き状態に戻す手段と、待ち行列から仮予約要求を取り出し、空き状態になった時間帯を仮予約された状態に変更し、仮予約要求の発行者を識別する情報を格納する手段と、前記待ち行列から取り出された仮予約要求の発行者に対して、仮予約された時間帯とそれに対応する仮予約 ID を含む通知を送る手段と、を備えた構成としてもよい。

【0036】

本発明において、ジョブの集合を資源に割り当てるジョブスケジュール装置であり、ジョブが使用する可能性のある資源のそれぞれについて、該資源を管理する資源管理装置に対して、仮予約要求を送る第 1 の手段と、仮予約された資源を対象として、各ジョブを割り当てる資源と、実行する時間帯とを決める第 2 の手段と、前記第 2 の手段によって決定された資源のそれぞれについて、該資源を管理する資源管理装置に対して、ジョブ予約要求を送る第 3 の手段と、前記第 1 の手段によって仮予約された資源のそれぞれについて、該資源を管理する資源管理装置に対して、仮予約取消要求を送る第 4 の手段と、を備えた構成としてもよい。

【0037】

本発明において、前記ジョブスケジューラにおいて、前記第1の手段で待ち行列に格納された仮予約要求がある場合に、一定時間が経過するか、待ち行列から取り出されたことを示す通知が到着してから、前記第2の手段に進む、構成としてもよい。

【0038】

本発明の他のアスペクトに係る方法は、1つ又は複数の資源を管理する資源管理装置による資源管理の方法であって、前記資源管理装置が、管理対象の資源を1つ又は複数指定してなる仮予約要求を受理した際に、前記管理対象の資源のうち空き状態である時間帯を仮予約された状態に変更して、有効期限を設けるとともに、前記仮予約要求の発行者を識別する情報を記憶部に格納する工程と、

前記資源管理装置が、前記管理対象の資源のうち、仮予約された時間帯が存在する資源についての待ち行列に、仮予約要求を格納する工程と、

前記資源管理装置が、前記仮予約された時間帯と、前記仮予約された時間帯に対応する仮予約識別情報（「仮予約ID」という）を、前記仮予約要求に対する応答とする工程と、

を含む。

【0039】

本発明のさらに他のアスペクトに係るシステムは、1つ又は複数の資源の管理を行う、少なくとも1つの資源管理装置と、少なくとも1つの前記資源管理装置と通信接続される、少なくとも1つのジョブスケジューラ装置と、を含む分散システムであって、前記ジョブスケジューラ装置が、ジョブのスケジューリングに先立って、前記ジョブで用いる資源の仮予約を行い、前記資源管理装置が、前記ジョブスケジューラ装置により発行される、仮予約とジョブ予約との2種の予約要求を受理し、前記仮予約は、資源の予約状況の取得と、資源のロックを組み合わせた操作であって、前記資源管理装置の管理下にある資源のいずれかを指定して実行される。前記資源管理装置は、前記ジョブスケジューラ装置より受け取った前記仮予約要求の対象の資源に対して、他の仮予約が存在せず、かつ、空き時

間が存在する場合には、当該時間帯を仮予約された状態に設定し、前記仮予約された時間帯を、前記仮予約要求に対する応答とする手段と、前記資源管理装置が、前記仮予約された時間帯に、有効期限と、該当する仮予約要求を発行したジョブスケジュール装置の情報とを記録する手段と、を有し、前記ジョブ予約は、指定されたジョブの実行のために資源を事前予約する操作であって、前記資源管理装置が管理する資源のうちの1つと、前記資源における時間帯を指定したジョブ予約要求として実行される。前記資源管理装置は、前記ジョブ予約要求で指定された時間帯がすでに仮予約された時間帯の一部であり、かつ、前記ジョブ予約要求の発行者と前記仮予約要求の発行者とが互いに一致する場合に、前記ジョブ予約要求を受理し、指定された時間帯をジョブ予約された状態とする手段と、ジョブ予約で指定された開始時刻に達すると、当該資源をジョブの実行のために割り当てる制御を行い、及び／又は、前記ジョブ予約で指定された終了時刻になると、当該の資源のジョブへの割り当てを停止する制御を行う手段と、を備えている。

【0040】

本発明のさらに他のアスペクトに係るプログラム（プログラム製品）は、1つ又は複数の資源を管理する資源管理装置を構成するコンピュータに、管理対象の資源を1つ又は複数指定してなる仮予約要求を受理した際に、前記管理対象の資源のうち、空き状態である時間帯を、仮予約された状態に変更して、有効期限を設けるとともに、前記仮予約要求の発行者を識別する情報を記憶部に格納する処理と、

前記管理対象の資源のうち、仮予約された時間帯が存在する資源についての待ち行列に、仮予約要求を格納する処理と、

前記仮予約された時間帯と、前記仮予約された時間帯に対応する仮予約識別情報（「仮予約ID」という）を応答とする処理と、

を実行させるプログラムよりなる。

【0041】

本発明によれば、資源の利用を予約するために、資源管理装置機構に対して、仮予約とジョブ予約の2段階が設けられている。仮予約は短時間のみ有効であり

、他の仮予約およびジョブ予約とは共存できず、資源を実際に利用する場合、仮予約した後で、ジョブ予約を実行するものとされ、すでに仮予約された資源に、仮予約を試みると待ち行列に入れ、後で仮予約が成功することを保証している。ユーザ端末は、ジョブスケジューラにジョブの実行を要求し、ジョブスケジューラは、資源管理装置に仮予約操作を実行した後で、ジョブを割り当てる資源を決定し、ジョブ予約を実行し、ジョブの実行に必要な資源を予約する。

【0042】

本発明によれば、同時に、複数の資源を使うジョブや依存関係のあるジョブ群の一部のみ予約に失敗するという状況が発生しない。このため、資源が分散して管理される分散システムにおいて、資源の利用効率を向上できる。

【0043】

[発明の原理]

まず本発明の原理を説明し、つづいて実施形態について説明する。本発明においては、複数の資源管理装置（図5の100）と、単一または複数のジョブスケジューラ（図5の200）から構成される。複数の資源（図5の300）は、資源管理装置の1つにより管理される。

【0044】

資源管理装置は、ジョブスケジューラにより発行される仮予約と、ジョブ予約という2種類の予約要求を受理する。また、これらの予約要求に対応して、仮予約取消要求およびジョブ取消要求の2種類の取消要求を受理する。

【0045】

仮予約は、資源の予約状況の取得と、資源のロックを組み合わせた操作であり、資源管理装置の管理下にある資源のどれかを指定して実行される。仮予約要求の対象の資源に他の仮予約が存在せず、かつ空き時間が存在すれば、その時間帯が仮予約された状態となり、仮予約された時間帯を応答とする。空き時間とは、仮予約とジョブ予約のいずれも存在しない時間帯である。仮予約された時間帯には、有効期限と、該当する仮予約要求を発行したジョブスケジューラの情報が記録される。有効期限に達した仮予約は自動的に破棄される。

【0046】

ジョブ予約は、特定のジョブの実行のために資源を事前予約する操作であり、資源管理装置が管理する資源のうちの1つと、その資源における時間帯を指定したジョブ予約要求として実行される。指定された時間帯がすでに仮予約された時間帯の一部であり、かつそのジョブ予約要求の発行者と仮予約要求の発行者が一致する場合にジョブ予約要求は受理され、指定された時間帯がジョブ予約された状態となる。資源管理装置は、ジョブ予約で指定された開始時刻に達すると、当該の資源をジョブの実行のために割り当て、ジョブ予約で指定された終了時刻になると当該の資源のジョブへの割り当てを停止する。

【0047】

本発明では、ユーザが投入するジョブは、以下の順序で実行される。まず、ユーザは、ユーザ端末（図5の400）を通して、ジョブスケジューラに、ジョブを投入する（図5（1））。

【0048】

ジョブスケジューラは、投入されたジョブを定期的に各資源に割り当てる処理を行う。ジョブスケジューラによる資源の割り当ては、資源管理装置に対する仮予約要求の送信（図5（2））と、仮予約に成功した資源を対象とするジョブの実行スケジュールの作成（図5（3））と、実行スケジュールに従ったジョブ予約要求の送信（図5（4））とからなる。資源管理装置は、ジョブスケジューラから受信したジョブ予約要求に従って、管理下の資源をジョブの実行のために割り当てる（図5（5））。

【0049】

あるジョブスケジューラによる仮予約が有効である間に別のジョブスケジューラが同一の資源を対象として仮予約要求を発行すると、後者のジョブスケジューラは仮予約を獲得できない。この場合、仮に前者のジョブスケジューラが当該の資源にジョブを割り当てない場合でも、後者のジョブスケジューラは当該の資源にジョブを割り当てることはできない。この問題を解決するため、資源管理装置は、すでに仮予約された資源を対象とする仮予約要求を受け取ると、その仮予約要求を資源ごとに設けられる待ち行列に格納するようにする。仮予約取消要求または有効期限切れのために仮予約が破棄されると、破棄された仮予約と同一の資

源を対象とする仮予約要求が待ち行列から取り出され、有効な仮予約となる。待ち行列から仮予約要求が取り出されると、資源管理装置がその仮予約要求を発行したジョブスケジューラに通知を送ることによって、ジョブスケジューラはそのことを知ることができる。

【0050】

ジョブスケジューラが発行した仮予約要求のどれかが待ち行列に入ると、ジョブスケジューラはそれらの仮予約が有効になるのを待ってから実行スケジュールの作成を開始する。ただし、一定時間内に仮予約要求が有効にならない場合は、その時点で待ち行列に残っている仮予約を無視して実行スケジュールの作成に進む。

【0051】

上記の構成を採用し、ジョブスケジューラがジョブのスケジューリングに先立って資源を仮予約することで、ジョブスケジューラは資源の予約状況を正確に把握することができる。このことから、資源の正確な予約状況が把握できないことによる予約の失敗を避けることができ、最適な資源の組み合わせを選ぶことが可能になる。これにより、本発明の目的を達成することができる。資源管理装置およびスケジューラにおける上記処理・機能は、それぞれを構成するコンピュータで実行されるプログラムにより実現することができる。

【0052】

【発明の実施の形態】

次に、本発明の実施の形態について図面を参照して詳細に説明する。図5は、本発明の一実施の形態のシステム構成を示す図である。図5を参照すると、本発明の一実施の形態は、複数の資源管理装置100と、複数のジョブスケジューラ200と、複数の資源300と、複数のユーザ端末400（図5では1台の端末のみが示されている）と、を備えて構成されている。ここで、資源管理装置100、ジョブスケジューラ200、およびユーザ端末400は、プログラム制御により動作するコンピュータを含む。資源300には、コンピュータ、記憶装置、ネットワーク機器などの各種装置が含まれる。ユーザ端末400はジョブスケジューラ200を内蔵していても良い。また、資源管理装置100は、任意の数の

資源 300 を内蔵していても良い。またユーザ端末 400 は 1 つであってもよい。本発明は、複数の資源管理装置 100、複数のジョブスケジューラ 200 をそれぞれ 1 つとしたシステム構成にも同様にして適用できる。

【0053】

図 6 は、図 5 に示した本発明の一実施の形態における資源管理装置 100 の構成を示す図である。図 6 を参照すると、資源管理装置 100 は、仮予約要求受信手段 110 と、ジョブ予約要求受信手段 120 と、仮予約取消要求受信手段 130 と、ジョブ予約取消要求受信手段 140 と、スケジュール管理手段 150 と、スケジュール記憶部 160 と、仮予約要求記憶部 170 とを備えて構成されている。

【0054】

スケジュール記憶部 160 には、資源管理装置の管理対象である資源のそれぞれについて、どの時間帯にどのジョブが予約されているかが記録される。

【0055】

仮予約要求記憶部 170 には、資源管理装置の管理対象である資源のそれぞれについて待ち行列が設けられる。この待ち行列には、その資源を対象とする待機中の仮予約要求が格納される。

【0056】

これらの手段は、それぞれ概略次のように動作する。仮予約要求受信手段 110 は、ジョブスケジューラ 200 からの仮予約要求を受信し、スケジュール記憶部 160 または仮予約要求記憶部 170 に仮予約についての情報を追加する。

【0057】

ジョブ予約要求受信手段 120 は、ジョブスケジューラ 200 からのジョブ予約要求を受信し、スケジュール記憶部 160 にジョブ予約の情報を追加する。ジョブ予約を行うためには、あらかじめ仮予約要求受信手段 110 を通してスケジュール記憶部 160 に仮予約が登録されている必要がある。

【0058】

仮予約取消要求受信手段 130 は、ジョブスケジューラ 200 からの仮予約取消要求を受信し、スケジュール記憶部 160 または仮予約要求記憶部 170 から

該当する仮予約の情報を破棄する。それに伴い、仮予約要求記憶部 170 に記録された待機中の仮予約要求をスケジュール記録部 160 に移動する場合がある。

【0059】

ジョブ予約取消要求受信手段 140 は、ジョブスケジューラ 200 からのジョブ予約取消要求を受信し、スケジュール記録部 160 から該当するジョブ予約の情報を破棄する処理と、ジョブに割り当てられた資源を解放する処理を行う。

【0060】

スケジュール管理手段 150 は、スケジュール記録部 160 に記録された予約ジョブおよび仮予約の情報を監視して、それらの情報に従った資源の割り当て・解放や仮予約の破棄などの処理を行う。図 6 に示した資源管理装置 100 の上記各手段は、資源管理装置 100 を構成するコンピュータにより、その機能を実現することができる。

【0061】

図 7 は、図 5 に示した本発明の一実施の形態におけるジョブスケジューラ 200 の構成を示す図である。図 7 を参照すると、ジョブスケジューラ 200 は、ジョブ実行要求受信手段 210 と、ジョブ取消要求受信手段 220 と、ジョブスケジューリング手段 230 と、待機ジョブ記憶部 240 と、スケジュール記憶部 250 とを備えて構成されている。待機ジョブ記憶部 240 には、ユーザ端末により投入されたジョブの情報が一時的に格納される。スケジュール記憶部 250 には、各資源ごとに、予約が確定したジョブの情報が格納される。

【0062】

これらの手段は、それぞれ概略次のように動作する。ジョブ実行要求受信手段 210 は、ユーザ端末 400 からのジョブ実行要求を受信し、その内容（ジョブ実行要求に含まれるジョブ情報等）を、待機ジョブ記憶部 240 に、追加（記録）する処理を行う。

【0063】

ジョブ取消要求受信手段 220 は、ユーザ端末 400 からのジョブ取消要求を受信し、待機ジョブ記憶部 240、またはスケジュール記憶部 250 に登録されている、該当するジョブを破棄するとともに、資源管理装置 100 に対して、ジ

ジョブ予約取消要求の送信を行う。

【0064】

ジョブスケジューリング手段230は、待機ジョブ記憶部240に記録されているジョブを、定期的に取り出し、それらのジョブを、資源に割り当てる処理を行う。ジョブの資源への割り当てには、資源管理装置100への仮予約要求の発行、ジョブ実行スケジュールの作成、および、資源管理装置100へのジョブ予約要求の発行が含まれる。ジョブスケジューラ200の上記各手段は、ジョブスケジューラ200を構成するコンピュータによりその機能を実現することができる。

【0065】

資源管理装置100とジョブスケジューラ200の上記した各手段間の相互作用の概略を、図8および図9を用いて説明する。図8には、ユーザ端末400がジョブスケジューラ200にジョブを投入してから、資源上でジョブの実行が完了するまでの処理の流れが、模式的に示されている。

【0066】

ステップ1：ユーザ端末400が発行したジョブ実行要求を、ジョブスケジューラ200のジョブ実行要求受信手段210が受信し、その内容を待機ジョブ記憶部240に格納する（図8（1））。

【0067】

ステップ2：ジョブスケジューラ200のジョブスケジューリング手段230が、待機ジョブ記憶部240からジョブの情報を読み出し、資源管理装置100の仮予約要求受信手段110に対して、仮予約要求を送信する。仮予約要求受信手段110は、スケジュール記憶部160に、仮予約の情報を追加する（図8（2））。

【0068】

ステップ3：ジョブスケジューラ200のジョブスケジューリング手段230が、ジョブ実行スケジュールの作成完了後に、資源管理装置100のジョブ予約要求受信手段120に対して、ジョブ予約要求を送信する。ジョブ予約要求受信手段120は、スケジュール記憶部160に、ジョブ予約の情報を追加する（図

8(3))。

【0069】

ステップ4：ジョブスケジューラ200のジョブスケジューリング手段230が、資源管理装置100の仮予約取消要求受信手段130に対して、仮予約取消要求を送信する。仮予約取消要求受信手段130は、スケジュール記憶部160から仮予約の情報を削除する（図8（4））。

【0070】

ステップ5：資源管理装置100のスケジュール管理手段150は、スケジュール記憶部160に記録されたジョブ予約の開始時刻になると、予約された資源300をジョブに割り当てる。同様に、ジョブ予約の終了時刻になると、ジョブに割り当てた資源300を解放する（図8（5））。

【0071】

図9には、上記手順における、ジョブスケジューラ200内の待機ジョブ記憶部240およびスケジュール記憶部250の状態の変化の一例が示されている。

【0072】

図9（A）は、上記ステップ1の直後の状態を示している。スケジュール記憶部250には、以前に予約したジョブの情報が格納されている。また、待機ジョブ記憶部240には、ステップ1で受け取ったジョブの情報が格納されている。

【0073】

図9（B）は、上記ステップ2の直後の状態を示している。ここではスケジュール記憶部250に、仮予約に成功した時間帯の情報が追加されている。図9（C）に示すように、同一の資源について不連続な時間帯について仮予約がなされることがあり得る。

【0074】

図9（C）は、上記ステップ3におけるジョブ予約要求送信の直前の状態を示している。スケジュール記憶部250にスケジューリングされたジョブの情報が追加されているが、この時点ではまだ予約は確定していない。

【0075】

図9 (D) は、上記ステップ4の直後の状態を示している。ジョブ予約要求の発行によって資源の予約が確定し、待機ジョブ記憶部240から該当するジョブの情報が削除されている。

【0076】

次に、本実施の形態の全体の動作について詳細に説明する。以降では、最初に資源管理装置100の動作について説明した後、ジョブスケジューラ200の動作について説明する。

【0077】

まず、資源管理装置100の各手段(図6参照)の動作を詳細に説明する。

【0078】

仮予約要求受信手段110は、ジョブスケジューラ200からの仮予約要求を受け取り、スケジュール記憶部160の内容を参照して、仮予約要求を処理する。

【0079】

図10は、資源管理装置100における仮予約要求受信手段110の動作手順を説明するための流れ図である。図6及び図12を参照して、仮予約要求受信手段110の動作を説明する。

【0080】

まず、ジョブスケジューラ200からの要求内容を受信する(ステップS111)。要求内容には、要求を発行するジョブスケジューラの識別子と、仮予約の対象とする資源の識別子が含まれる。

【0081】

次に、受信した仮予約要求に対応する新しい仮予約IDを生成する(ステップS112)。以後、この仮予約IDを用いて、仮予約および仮予約要求が識別される。

【0082】

次に、スケジュール記憶部160を参照して、ステップS111で指定された資源がすでに他のジョブスケジューラ200によって仮予約されているか否かを調べる(ステップS113)。仮予約されている場合には(ステップS113で

のYES分岐)、ステップS115に、仮予約されていない場合には(ステップS113でのNO分岐)、ステップS114に進む。

【0083】

ステップS114では、スケジュール記憶部160に、新しい仮予約の情報を追加する。仮予約の情報には、仮予約される時間帯と、仮予約IDと、仮予約要求の発行者の識別子が含まれる。仮予約される時間帯は、対象とする資源の全時間帯のうち、ジョブ予約がなされている時間帯を除いた部分である。また、システム管理上の方針により、仮予約の対象とする時間帯を制限することもできる。スケジュール記憶部160に追加された仮予約には、有効期限が付加される。有効期限をどのように設定するかは、システム管理上の方針により決めることができる。

【0084】

ステップS115では、受信した仮予約要求を待機状態として仮予約要求記憶部170に格納する。

【0085】

ここで、仮予約が、スケジュール記憶部160に登録される場合と、仮予約要求記憶部170に登録される場合との違いについて具体例を用いて詳細に説明する。図11(A)乃至図11(D)には、同一の資源について、スケジュール記憶部160中のスケジュールと、仮予約要求記憶部170の待ち行列の状態が示されている。なお、この実施形態で、待ち行列はFIFO(先入れ先出し)型としており、キューエレメントは、キューのテイルに挿入され、キューのヘッド(先頭)から取り出される。なお、本発明において、待ち行列はかかる構成にのみ限定されるものでない。

【0086】

図11(A)には、ある資源に関するスケジュールとして、ジョブ1とジョブ2の2つのジョブ予約がなされており、仮予約要求記憶部170中の同じ資源についての待ち行列は空であるという状況が示されている。

【0087】

図11(B)では、仮予約が存在しないため、図11(A)の状態において仮

予約要求を受信すると、図10のステップS114に進み、スケジュールの中でジョブ1およびジョブ2で予約されている時間帯以外の時間帯が仮予約された状態となる（図11（B）の「仮予約3」参照）。

【0088】

図11（B）の状態において、新たな仮予約要求を受信すると、すでに仮予約が存在するために、図10のステップS115に進んで待ち行列に仮予約要求が追加される。

【0089】

図11（C）には、図11（B）の状態から「仮予約要求4」、「仮予約要求5」の順で仮予約要求が到着した場合の状況が示されている。

【0090】

図11（C）の状態の後で、「仮予約3」が有効期限切れ、または仮予約取消要求によって破棄された場合、待ち行列の先頭にある「仮予約要求4」がスケジュールに登録される（図11（D））。

【0091】

なお、待ち行列からスケジュールへの仮予約の移動を行う処理については、後述の仮予約取消要求受信手段130およびスケジュール管理手段150の動作手順で説明する。

【0092】

次に、資源管理装置100におけるジョブ予約要求受信部120の動作を詳細に説明する。

【0093】

ジョブ予約受信部120は、ジョブスケジューラ200からのジョブ予約要求を受け取り、スケジュール記憶部160にジョブ予約を登録する処理を行う。ジョブスケジューラから受け取るジョブ予約要求には、要求を発行するジョブスケジューラの識別子と、ジョブを実行するユーザの識別子と、予約の対象とする資源の識別子と、予約する時間帯の開始時刻及び終了時刻と、ジョブ予約の識別に用いるジョブIDが含まれる。

【0094】

なお、予約対象の資源の識別子については、以前に発行された仮予約IDによって指定するようにしても良い。また、ジョブ予約要求にジョブIDを含めるのではなく、ジョブ予約要求受信部120によって新しいジョブIDを発行するようにしても良い。

【0095】

ジョブ予約要求が成功するためには、予約の対象である時間帯が仮予約された時間帯に含まれており、かつその仮予約に対応する仮予約要求の発行者がジョブ予約要求の発行者と一致している必要がある。この条件を満たしている場合、受信したジョブ予約要求の内容がスケジュール記憶部160に登録される。

【0096】

次に、図12は、資源管理装置100における仮予約取消要求受信手段130の動作を説明するための流れ図である。図6及び図12を参照して、仮予約取消要求受信手段130の動作を説明する。仮予約取消要求受信手段130は、ジョブスケジューラ200からの仮予約取消要求を受信し、スケジュール記憶部160に記録された仮予約または仮予約要求記憶部170に記録された仮予約要求を破棄する処理を行う。

【0097】

まず、仮予約取消要求受信手段130は、要求内容を受信する（ステップS131）。要求内容には、要求を発行するジョブスケジューラの識別子と、取り消しの対象となる仮予約／仮予約要求を識別する仮予約IDが含まれる。もし該当する仮予約／仮予約要求が存在しない場合、あるいは該当する仮予約要求を発行したジョブスケジューラと仮予約取消要求を発行したジョブスケジューラが異なる場合には、仮予約取消要求は失敗する。

【0098】

次に、ステップS131で受信した仮予約IDが、仮予約要求記憶部170中の待ち行列の要素に該当するかどうかを判別する（ステップS132）。判定の結果、受信した仮予約IDが待ち行列中の仮予約要求に該当する場合（ステップS132のYES分岐）、待ち行列から該当する要素を削除して処理が完了する（ステップS133）。そうでない場合（ステップS132のNO分岐）、すな

わち仮予約 ID がスケジュール記憶部 160 に記録された仮予約に該当する場合、ステップ S 134 以降の処理を行う。ステップ S 134 では、スケジュール記憶部 160 中の仮予約 ID に該当する仮予約を削除する。

【0099】

次に、仮予約要求記憶部 170 中に、ステップ S 134 で削除した仮予約の対象の資源についての待ち行列に要素があるか否かを調べる（ステップ S 135）。該判定の結果、待ち行列が空であれば（ステップ S 135 の NO 分岐）、処理を完了する。待ち行列が空でない場合（ステップ S 135 の YES 分岐）、ステップ S 136 に進む。

【0100】

ステップ S 136 では、待ち行列の先頭要素をスケジュール記憶部 160 に移動する処理を行う（図 11（C）－図 11（D）参照）。

【0101】

仮予約される時間帯の決定方法は、前記ステップ S 114 と同様である。それに加え、移動される仮予約要求を発行したジョブスケジューラ 200 に対して、待機中の仮予約要求が実行されたことを通知する。以降では、この通知を「仮予約成功通知」と呼ぶ。

【0102】

次に、資源管理装置 100 におけるジョブ予約取消要求受信手段 140 の動作を詳細に説明する。ジョブ予約取消要求受信手段 140 は、ジョブスケジューラ 200 からジョブ予約取消要求を受信し、該当するジョブを破棄する処理を行う。ジョブ予約取消要求には、要求を発行するジョブスケジューラの識別子と、取り消しの対象とするジョブ予約のジョブ ID が含まれる。

【0103】

ジョブ取消処理が成功するためには、ジョブ予約取消要求に含まれるジョブ ID に対応ジョブ予約がスケジュール記憶部 160 に登録されており、かつ、そのジョブ予約要求の発行者と、ジョブ予約取消要求の発行者とが互いに一致している必要がある。この条件が満たされていれば、スケジュール記憶部 160 から該当するジョブの情報を削除する。さらに、該当するジョブにすでに資源が割り当

てられている場合、当該ジョブが使用する資源を解放する処理を行う。資源を解放する処理は、後述のスケジュール管理手段による資源の解放処理と同様である。

【0104】

次に、資源管理装置100におけるスケジュール管理手段150の動作を詳細に説明する。スケジュール管理手段150は、スケジュール記憶部160に記録された予約情報に従って、以下の処理を行う。

【0105】

- ・ 予約開始時刻に達したジョブについて、資源の割り当て処理を行う。

【0106】

- ・ 予約終了時刻に達したジョブについて、資源の解放処理を行う。

【0107】

- ・ 有効期限に達した仮予約を破棄する。

【0108】

上記資源の割り当て処理とは、予約された資源を実際に利用可能な状態にすることを指し、実際に行われる処理は管理対象の資源の種類によって異なる。例えば、資源がコンピュータである場合にはユーザプログラムの起動、記憶装置であればリモートファイルのオープン、ネットワーク機器であれば帯域の割り当てなどが該当する。

【0109】

上記資源の解放処理とは、資源の割り当て処理によってジョブに割り当てていた資源を利用不可能な状態に戻し、他のジョブへの割り当てを可能にすることを指す。資源の解放処理の具体例としては、例えばユーザプログラムの停止、リモートファイルのクローズ、帯域の割り当て停止などが該当する。

【0110】

上記仮予約の破棄は、スケジュール記憶部160に記録された仮予約のうち、有効期限に達したものを削除する処理である。さらに、破棄される仮予約の対象である資源についての、空でない待ち行列が、仮予約要求記憶部170中に存在する場合、その先頭要素である仮予約要求をスケジュール記憶部160に登録し

、ジョブスケジューラ 200 に対して、仮予約成功通知を送る。この部分の処理は、図 12 の前記ステップ S 136 と共通である。

【0111】

次に、ジョブスケジューラ 200 の各手段の動作を詳細に説明する。まず、ジョブ実行要求受信手段 210 の動作を詳細に説明する。

【0112】

ジョブ実行要求受信手段 210 は、ユーザ端末 400 からのジョブ実行要求を受信し、待機ジョブ記憶部 240 に追加する処理を行う。ジョブ実行要求には、要求を発行するユーザの識別子と、使用する資源の記述が含まれる。使用する資源の記述には、コンピュータや記憶装置などの資源の種類、CPU 性能などの性能に関する要請、複数の資源を割り当てる際の依存関係が含まれる。ジョブ実行要求を受信すると、ジョブ実行要求受信手段は新しいジョブ ID を生成して、受信したジョブ実行要求の内容とともに待機ジョブ記憶部 240 に追加する。ここで追加されるジョブは待機状態である。さらに、生成されたジョブ ID を、ジョブ実行要求を発行したユーザ端末に通知する。

【0113】

次に、図 13 は、ジョブスケジューラ 200 のジョブ取消要求受信手段 220 の動作を説明するための流れ図である。図 7 及び図 13 を参照して、ジョブ取消要求受信手段 220 の動作を説明する。まず、ユーザ端末から要求内容を受信する（ステップ S 221）。要求内容には、ジョブ取消要求を発行するユーザの識別子と、取り消しの対象であるジョブ ID が含まれる。ジョブ取消要求を発行するユーザと当該ジョブに対応するジョブ実行要求を発行したユーザが一致しない場合には、ジョブ取消要求を拒否する。

【0114】

次に、待機ジョブ記憶部 240 およびスケジュール記憶部 250 を参照して、ステップ S 221 で受信したジョブ ID に該当するジョブの情報がいずれに含まれるかを調べる（ステップ S 222）。判定の結果、待機ジョブ記憶部 240 に含まれる場合には（ステップ S 222 の YES 分岐）、ステップ S 223 を実行し、スケジュール記憶部 250 に含まれる場合には（ステップ S 222 の NO 分

岐)、ステップS 2 2 4 およびS 2 2 5 を実行する。いずれにも含まれない場合にはジョブ取消要求は失敗する。

【0115】

ステップS 2 2 3 では、待機ジョブ記憶部 2 4 0 から、当該ジョブの情報を削除して処理を完了する。

【0116】

一方、ステップS 2 2 4 では、当該ジョブを予約した資源管理装置 1 0 0 に対してジョブ予約取消要求を送る。当該ジョブが予約されている資源がどれであるかはスケジュール記憶部 2 5 0 から取得する。

【0117】

ステップS 2 2 5 では、スケジュール記憶部 2 5 0 から当該ジョブを削除して処理を完了する。

【0118】

次に、図 1 4 は、ジョブスケジューラ 2 0 0 のジョブスケジューリング手段 2 3 0 の動作を説明するための流れ図である。図 7 及び図 1 4 を参照して、ジョブスケジューリング手段 2 3 0 の動作を説明する。ジョブスケジューリング手段 2 3 0 は、定期的に待機ジョブ記憶部 2 4 0 の内容を監視して、待機状態のジョブが存在すれば以下に説明するスケジューリング処理を行う。

【0119】

まず、待機ジョブ記憶部 2 4 0 に格納されている待機中のジョブのうち、スケジューリングの対象となるジョブを先着順に取り出す(ステップS 2 3 1)。ここで取り出すジョブの数は、システム管理上の方針により指定する。

【0120】

次に、ステップS 2 3 1 で取り出した待機中のジョブを割り当てることが可能な各資源について、当該資源を管理する資源管理装置 1 0 0 に対して仮予約要求を発行する(ステップS 2 3 2)。

【0121】

次に、ステップS 2 3 2 において仮予約要求記憶部 1 7 0 に格納された仮予約要求があるかどうかを調べる(ステップS 2 3 3)。仮予約要求がある場合には

(ステップ S 2 3 3 の Y E S 分岐)、ステップ S 2 3 4 に進み、仮予約要求がない場合には (ステップ S 2 3 3 の N O 分岐)、ステップ S 2 3 5 に進む。

【 0 1 2 2 】

ステップ S 2 3 4 では、一定時間待機し、資源管理装置 1 0 0 からの仮予約成功通知が到着するのを待つ。ここで待機する時間は、システム管理上の方針により決定されるが、少なくとも仮予約の有効期間よりも短い必要がある。一定時間が経過するか、あるいは待機状態のすべての仮予約要求についての仮予約成功通知が到着した場合、ステップ S 2 3 5 に進む。

【 0 1 2 3 】

ステップ S 2 3 5 では、仮予約に成功した資源を対象として、ジョブの実行スケジュールを作成する。実行スケジュールの作成とは、各ジョブをどの資源のどの時間帯に割り当てるかを決定する処理をいう。

【 0 1 2 4 】

この処理には、例えば上記非特許文献 2 等に記載された既存のジョブスケジューリングアルゴリズムを適用することができる。

【 0 1 2 5 】

上記非特許文献 2 で提案されているアルゴリズム D (min-minとして知られている) を用いることができる。なお、ステップ S 2 3 2 において仮予約要求記憶部 1 7 0 に格納され、かつ仮予約成功通知を受け取っていない仮予約要求がある場合、その仮予約要求の対象の資源はスケジューリングの対象に含めない。

【 0 1 2 6 】

次に、ステップ S 2 3 5 で作成した実行スケジュールに従って、各資源管理装置 1 0 0 に対してジョブ予約要求を発行する (ステップ S 2 3 6)。ここでジョブ予約に成功したジョブは、待機ジョブ記憶部 2 4 0 からスケジュール記憶部 2 5 0 に移される。

【 0 1 2 7 】

最後に、ステップ S 2 3 2 で仮予約要求を送った各資源について、仮予約取消要求を資源管理装置 1 0 0 に対して発行する (ステップ S 2 3 7)。

【 0 1 2 8 】

仮予約取消要求の対象とする資源には、仮予約要求記憶部 170 に格納された仮予約の対象の資源も含まれる。

【0129】

【実施例】

上記実施の形態についてさらに詳細に説明すべく、次に、2つの具体的な実施例について説明する。第1の実施例は、「グリッドデータファーム」と呼ばれる種類の分散システムに本発明を適用した実施例である。

【0130】

図15に示すように、第1の実施例のシステムは、インターネット (The Internet) を介して接続される2つのドメインからなり、各ドメイン内の構成要素はローカルエリアネットワーク (LAN) で接続されている。ドメイン1内の構成要素には、ジョブスケジューラ (1) 200と、資源管理装置 (1) 100と、ユーザ端末 (1) 400と、ファイルサーバノード (以下、単に「ノード」と呼ぶ) (11) 300～ノード (14) 300とがある。資源管理装置 (1) 300は、ノード (11) 300～ノード (14) 300を管理対象とする。ドメイン2内の構成要素には、ジョブスケジューラ (2) 200と、資源管理装置 (2) 100と、ユーザ端末 (2) 400と、ノード (21) 300～ノード (24) 300とがある。資源管理装置 (2) 100は、ノード (21) 300～ノード (24) 300を管理対象とする。各ファイルサーバノードはディスク装置を内蔵するコンピュータであり、ファイルを複数の断片に分けて格納する分散ファイルサーバを構成する。さらに、ファイルサーバノードは計算ノードとしても機能し、自ノードが格納するファイル断片を処理対象として並列にジョブを実行する。

【0131】

図16に、各ノード内に格納されているファイル断片の構成を示す。例えば、図15のノード (11) 300は、ファイルAの断片0 (ファイルA#0) とファイルBの断片0 (ファイルB#0) を格納している。図16では、A、B、Cの3種類のファイルがあり、ファイルAは、#0から#3までの4つの断片からなり、ファイルBおよびファイルCは、#0と#1の2つの断片からなる。

【0132】

上記の構成を持つシステムにおいて、以下の2つのジョブが実行される場合の動作手順を説明する。

【0133】

・ジョブ1：ファイルAを処理対象とする。図15のユーザ（1）が、ユーザ端末（1）400を通して、ジョブスケジューラ（1）200にジョブを投入する。

【0134】

・ジョブ2：ファイルBを処理対象とする。ユーザ（2）が、ユーザ端末（2）400を通して、ジョブスケジューラ（2）200にジョブを投入する。

【0135】

ジョブ1とジョブ2のいずれも、4つのノードを使用する並列ジョブとして実行される。従って、ジョブの実行に関わるすべてのノードについて、同じ時間帯に予約する必要がある。

【0136】

各ジョブスケジューラにジョブが投入されると、一旦、待機ジョブ記憶部（図7の240）にジョブが格納された後、ジョブスケジューリング手段（図7の230）によって資源の割り当てが行われる。

【0137】

ジョブスケジューラ（2）200は、ジョブ2の実行に用いる可能性のある資源について仮予約要求を発行する。ジョブ2は、ファイルBを処理対象としており、ファイルBの断片を保持するノードは、図16から、ノード（11）、ノード（12）、ノード（21）、ノード（22）の4つであるので、これら4つのノードを対象として仮予約要求を発行する。この時点では他の仮予約が存在していないので、これらのノードについてすべての時間帯で仮予約が成功する。ジョブスケジューラ（2）200は、ジョブ2についての実行スケジュールの作成を開始する。

【0138】

次に、ジョブスケジューラ（1）200は、ジョブ1の実行に用いる可能性の

ある資源について仮予約要求を発行する。ジョブ1は、ファイルAを処理対象としており、ファイルAの断片を保持するノードは、図16から、ノード(11)～ノード(14)およびノード(21)～ノード(24)の8つである。したがって、これらのノードに関して仮予約要求を発行する。この時点では、ノード(11)、ノード(12)、ノード(21)、ノード(22)がジョブスケジューラ(2)200により仮予約されているため、残りのノード(13)、ノード(14)、ノード(23)、ノード(24)のみが仮予約に成功し、残りの仮予約要求については待ち行列に格納される。このため、一定時間経過するか、仮予約成功通知が到着するまではスケジューリング処理を中断する。この時点における、各資源管理装置のスケジュール記憶部160および仮予約要求記憶部170の内容を、図17に示す。

【0139】

次に、ジョブスケジューラ(2)200は、ジョブ2についての実行スケジュールの作成を完了して、資源管理装置1(100)、資源管理装置2(100)の各々に対してジョブ予約要求を発行し、続いて、仮予約取消要求を発行する。その結果、ジョブ2のノード(11)300およびノード(12)300への割り当てが確定し、ノード11、12、21、22についての仮予約が破棄される。さらに、ノード11、12、21、22についての待ち行列から、ジョブスケジューラ(1)が発行した仮予約要求が取り出され、ノード11、12、21、22がジョブスケジューラ(1)によって仮予約された状態になる。この時点における各資源管理装置のスケジュール記憶部160および仮予約要求記憶部170の内容を、図18に示す。

【0140】

次に、ジョブスケジューラ(1)200は、資源管理装置(1)100および資源管理装置(2)100からの仮予約成功通知を受け取り、ジョブの実行スケジュールの作成を開始する。ここでは、実行開始までの時間が短く、かつすべての同一のドメイン内にあるノード(21)、ノード(22)、ノード(23)、ノード(24)の4つのノードを選択し、これらのノードを対象としてジョブ予約要求を送る。ジョブ予約の完了後、資源管理装置(1)100および資源管理

装置 (2) 100 に対して、仮予約取消要求を送る。

【0141】

図 19 に、最終的なスケジューリング結果を示す。この例では、2 つの仮予約要求が衝突しているにもかかわらず、要求の失敗が発生することなくジョブスケジューリングが正常に完了している。

【0142】

以上の動作手順は、あくまでもその一例を示したものであり、ジョブの投入順序や通信遅延などの影響によって異なるスケジューリング結果になる可能性がある。その場合でも問題なくスケジューリングが完了する。

【0143】

一方、本実施例と同様のシステムにおいて、仮予約の機構が存在しない場合を考える。ジョブスケジューラ (2) 200 によるジョブ予約が先に実行されると、ノード (11) とノード (12) にジョブ 2 が予約される。その後、ジョブスケジューラ (1) 200 によってジョブ 1 についてのジョブ予約が行われるが、もしジョブスケジューラ (1) 200 がジョブ 2 の情報を知らないとする、ノード (11) ~ ノード (14) を対象としてジョブ予約が試みられる可能性がある。この場合、ノード (13) とノード (14) についてのみ予約が成功し、他の 2 つのノードを別途探す必要がある。

【0144】

その結果、図 20 に示すスケジューリング結果となる。ジョブ 1 は、WAN をまたいだ資源に割り当てられるため、図 19 のスケジューリング結果よりも、資源の利用効率が低くなる。

【0145】

次に本発明の第 2 の実施例について説明する。本発明の第 2 の実施例は、依存関係のある 2 つのジョブを実行する場合の例である。図 21 に示すように、第 2 の実施例のシステムは、3 つのドメインから構成され、ドメイン間は WAN によって接続されている。各ドメインの内部は LAN で接続されている。ドメイン 1 内の LAN には、資源管理装置 (1) 100 と、管理対象の資源 300 として、ルータ (1) 300 と、ノード (11) 300 ~ ノード (14) 300 の 4 つの

計算ノード、および、記憶装置（１）３００が接続される。ドメイン２内のＬＡＮには、資源管理装置（２）１００と、管理対象の資源３００として、ルータ（２）３００と、ノード（２１）３００～ノード（２４）３００の４つの計算ノード、および記憶装置（２）３００が接続される。ドメイン３内のＬＡＮには、資源管理装置（３）３００と、管理対象の資源としてルータ（３）３００と、ユーザ端末４００、およびジョブスケジューラ２００が接続される。各ルータ、計算ノード、およびファイルサーバは、同じドメインにある資源管理装置により管理される資源である。各ドメイン１、２、３のルータ１、２、３は、帯域予約機能を有し、特定のジョブによる通信のために必要な帯域を予約することができる。資源管理装置１００に対してネットワーク帯域を予約すると、その資源管理装置１００は該当するルータに対して帯域予約操作を行う。

【０１４６】

また、ドメイン１と２の記憶装置１、２は、ネットワーク（ＬＡＮまたはＷＡＮ）を介してデータの読み書きが可能なネットワークストレージである。

【０１４７】

上記の構成を有するシステムにおいて、以下の２つのジョブをユーザ端末を通して実行することを考える。

【０１４８】

・ジョブ１：８つの計算ノードを使用して計算を行い、その結果を任意の記憶装置上のファイルＡに書き込む。記憶装置への計算結果の書き込みには、１０Ｍｂｐｓのネットワーク帯域を要する。

【０１４９】

・ジョブ２：記憶装置上のファイルＡからデータを読み、その内容を２つの計算ノードを使用して可視化し、その結果を画像データとしてユーザ端末に送る。ファイルＡの読み込みには１０Ｍｂｐｓのネットワーク帯域を、画像データの送信には１Ｍｂｐｓのネットワーク帯域を要する。

【０１５０】

以下に、図２２を用いて、上記のジョブを投入したときの動作手順を示す。図２２は、各資源管理装置が保持するスケジュール記憶部１６０の内容の変化を示

している。

【0151】

まず、図22 (A) は、スケジューリング開始前の状態を示している。この状態では、ノード11、ノード12、ノード22、ノード23上に既存のジョブ予約が存在する。

【0152】

ドメイン3内のジョブスケジューラ200は、ユーザ端末400からジョブ1およびジョブ2についてのジョブ実行要求を受け取った後、ジョブスケジューリング手段230 (図7参照) による割り当て処理を開始する。最初に、ジョブスケジューラ200は、ドメイン1、2、3の資源管理装置1、2、3が管理する資源について仮予約要求を発行する。

【0153】

図22 (B) は、ジョブスケジューラ200による仮予約が完了した後の各資源管理装置1、2、3のスケジュール記憶部160の内容を示している。

【0154】

次に、ジョブスケジューラ200は、ジョブ1およびジョブ2を割り当てる資源と時間帯を選択する。ここで、ジョブ1におけるファイルAへの書き込みはジョブ2におけるファイルAの読み出しよりも先行する必要があるという制約と、ジョブ2におけるファイルAの読み出しに用いる記憶装置はジョブ1におけるファイルAの書き込みに用いた記憶装置と同一である必要であるという制約がある。

【0155】

作成した実行スケジュールに従って、ドメイン1、2、3の各資源管理装置に、ジョブ予約要求を発行した結果、図22 (C) に示す予約状況になる。

【0156】

ジョブ予約に成功すると、ジョブスケジューラ200は、ドメイン1、2、3の各資源管理装置に、仮予約取消要求を送り、仮予約を破棄する。

【0157】

ドメイン1、2、3の各資源管理装置における最終的なスケジュール記憶部1

60の内容が図23に示されている。すなわち、図23を参照すると、ドメイン1の資源管理装置(1)100のスケジュール記憶部160において、ノード(11)～ノード(14)でのジョブ1の計算処理、ルータ(1)でのジョブ1の帯域予約、記憶装置(1)におけるジョブ1のファイルの書き込みと、対応する時間帯において、ドメイン2の資源管理装置(2)100のスケジュール記憶部160において、ノード(21)～ノード(24)でのジョブ1の計算処理、ルータ(2)でのジョブ1の帯域予約が、それぞれ登録され、つづいて、ドメイン1において、ノード(11)～ノード(14)でのジョブ2の計算処理、ルータ(1)でのジョブ2の帯域予約、記憶装置(1)におけるジョブ2のファイル読み出しが登録され、対応する時間帯において、ドメイン3の資源管理装置(3)100のスケジュール記憶部160において、ジョブ2の帯域予約が登録されている。なお、上記した実行スケジュールは一例であり、必ずしもこの通りになるとは限らない。

【0158】

以上本発明を上記各実施例に即して説明したが、本発明は、上記実施例の構成にのみ限定されるものでなく、本発明の範囲内で当業者であればなし得るであろう各種変形、修正を含むことは勿論である。

【0159】

【発明の効果】

以上説明したように、本発明によれば、下記記載の効果を奏する。

【0160】

本発明の第1の効果は、資源が分散して管理される分散システムにおいて、資源の利用効率を向上できることにある。

【0161】

その理由は、本発明においては、同時に、複数の資源を使うジョブや依存関係のあるジョブ群の一部のみ予約に失敗するという状況が発生しないためである。

【0162】

ジョブに必要な資源を事前に予約する場合、各資源についての予約状況の情報をあらかじめ取得する必要がある。前記したとおり、従来の手法では、取得した

予約状況の情報と予約操作をする際の実際の予約状況が一致しない可能性があった。このため、同時に複数の資源を使うジョブや依存関係のあるジョブ群の一部のみ予約に失敗し、その結果可能な範囲で最適な資源の組み合わせが得られないことがある。これに対して、本発明によれば、仮予約の機構により、資源の予約状況の取得と、資源のロックとを同時に行うため、予約可能な資源を仮予約してから最適な資源の組み合わせを選ぶことができる。

【0163】

本発明の別の効果は、多数のジョブを実行する際の計算機およびネットワークの負荷の増大を抑止・低減することができる、ということである。

【0164】

その理由は、本発明においては、競合する仮予約要求を待ち行列に格納することによって、必要な資源が獲得できるまで仮予約要求を繰り返す必要がないからである。

【図面の簡単な説明】

【図1】

資源の事前予約の概念を説明するための図である。

【図2】

従来技術1のシステム構成を示す図である。

【図3】

従来技術2のシステム構成を示す図である。

【図4】

従来技術2を適用した場合に非効率な資源の割り当てが行われる例を示すための図である。

【図5】

本発明の実施の形態のシステム構成を示す図である。

【図6】

本発明の実施の形態における資源管理装置100の構成を示すブロック図である。

【図7】

本発明の実施の形態におけるジョブスケジューラ 200 の構成を示すブロック図である。

【図 8】

本発明の実施の形態における各手段間の相互作用を説明するための図である。

【図 9】

ジョブスケジューラにおけるスケジュール記憶部 250 の状態変化を示す図である。

【図 10】

資源管理装置の仮予約要求受信手段 110 の動作手順を示すフローチャートである。

【図 11】

資源管理装置の仮予約要求受信手段 110 における待ち行列の扱いを説明するための図である。

【図 12】

資源管理装置の仮予約取消要求受信手段 130 の動作手順を示すフローチャートである。

【図 13】

ジョブスケジューラのジョブ取消要求受信手段 220 の動作手順を示すフローチャートである。

【図 14】

ジョブスケジューラのジョブスケジューリング手段 230 の動作手順を示すフローチャートである。

【図 15】

本発明の第 1 の実施例のシステム構成を示す図である。

【図 16】

本発明の第 1 の実施例におけるファイルサーバノードの構成を示す図である。

【図 17】

本発明の第 1 の実施例におけるジョブスケジューリングの途中経過を示す図である。

【図 18】

本発明の第 1 の実施例におけるジョブスケジューリングの途中経過を示す図である。

【図 19】

本発明の第 1 の実施例におけるジョブスケジューリングの結果を示す図である。

【図 20】

本発明の第 1 の実施例において、仮予約の機構を用いない場合のジョブスケジューリングの結果を示す図である。

【図 21】

本発明の第 2 の実施例のシステム構成を示す図である。

【図 22】

本発明の第 2 の実施例におけるジョブスケジューリングの途中経過を示す図である。

【図 23】

本発明の第 2 の実施例におけるジョブスケジューリングの結果を示す図である。

【符号の説明】

- 100 資源管理装置
- 110 仮予約要求受信手段
- 120 ジョブ予約要求受信手段
- 130 仮予約取消要求受信手段
- 140 ジョブ予約取消要求受信手段
- 150 スケジュール管理手段
- 160 スケジュール記憶部
- 170 仮予約要求記憶部
- 200 ジョブスケジューラ
- 210 ジョブ実行要求受信手段
- 220 ジョブ取消要求受信手段

2 3 0 ジョブスケジューリング手段

2 4 0 待機ジョブ記憶部

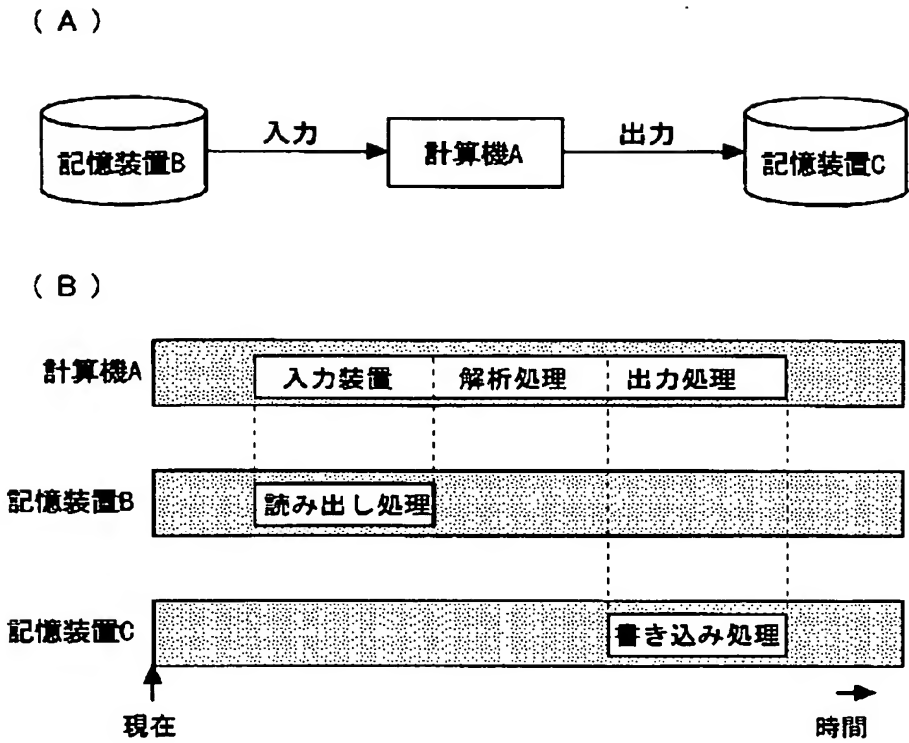
2 5 0 スケジュール記憶部

3 0 0 資源

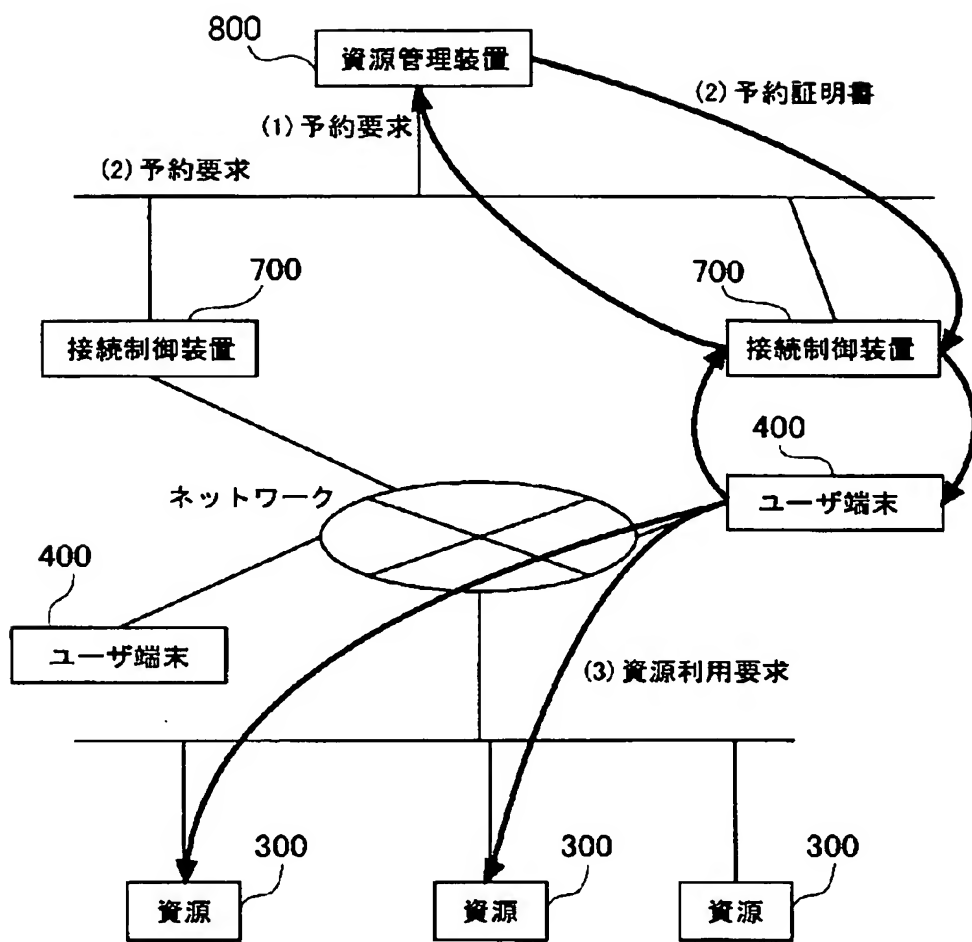
4 0 0 ユーザ端末

【書類名】 図面

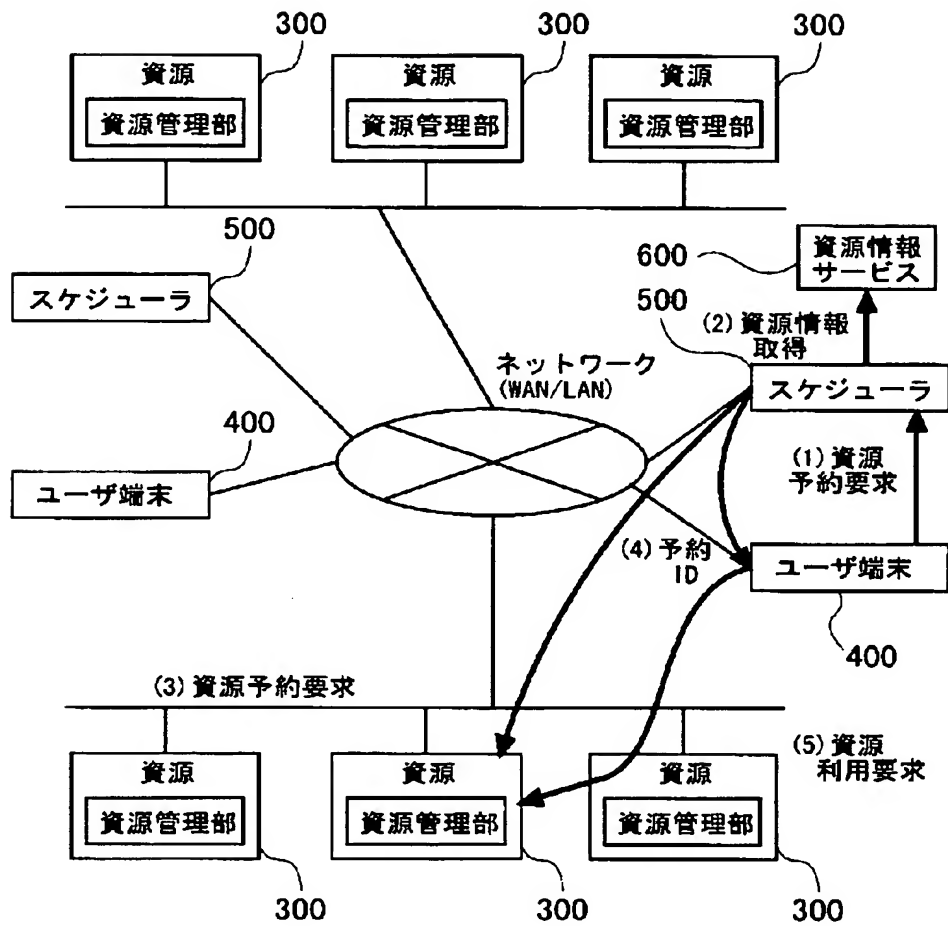
【図 1】



【圖 2】

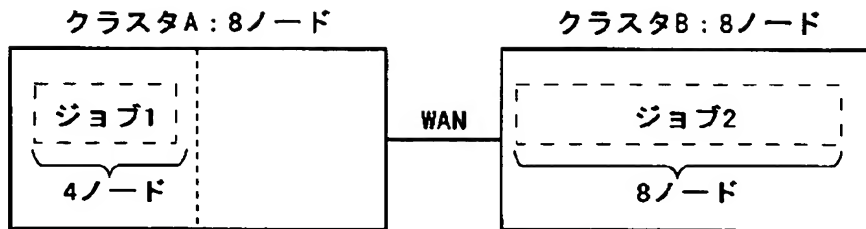


【図 3】



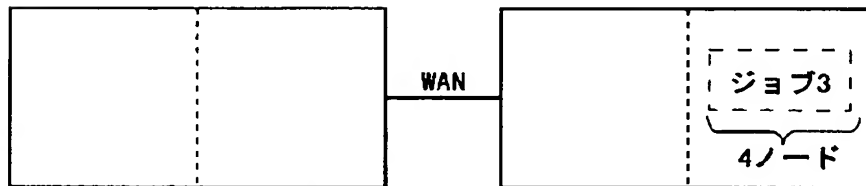
【図 4】

(A)



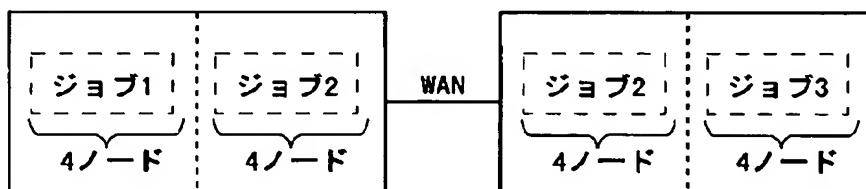
ジョブの割り当て(予約要求発行前)

(B)



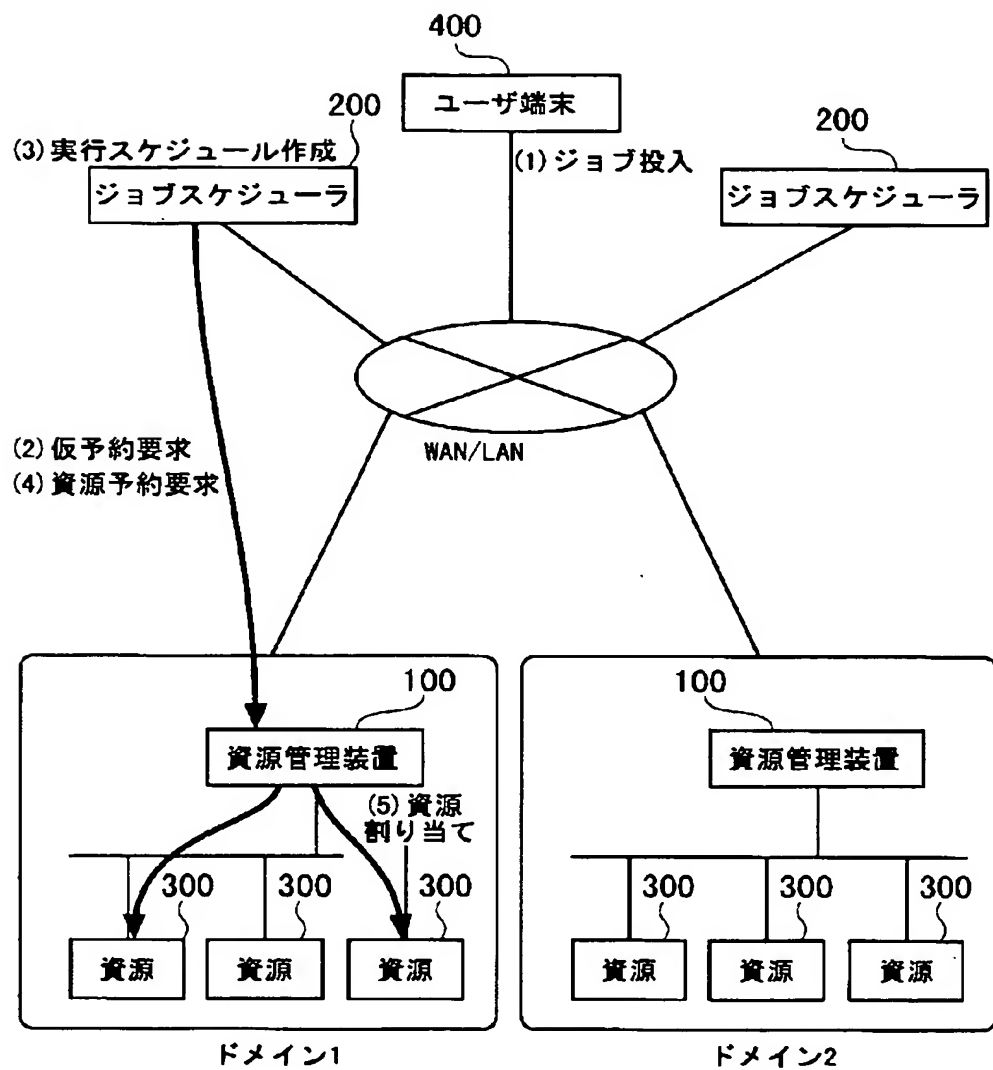
予約要求発行時点での予約状況

(C)

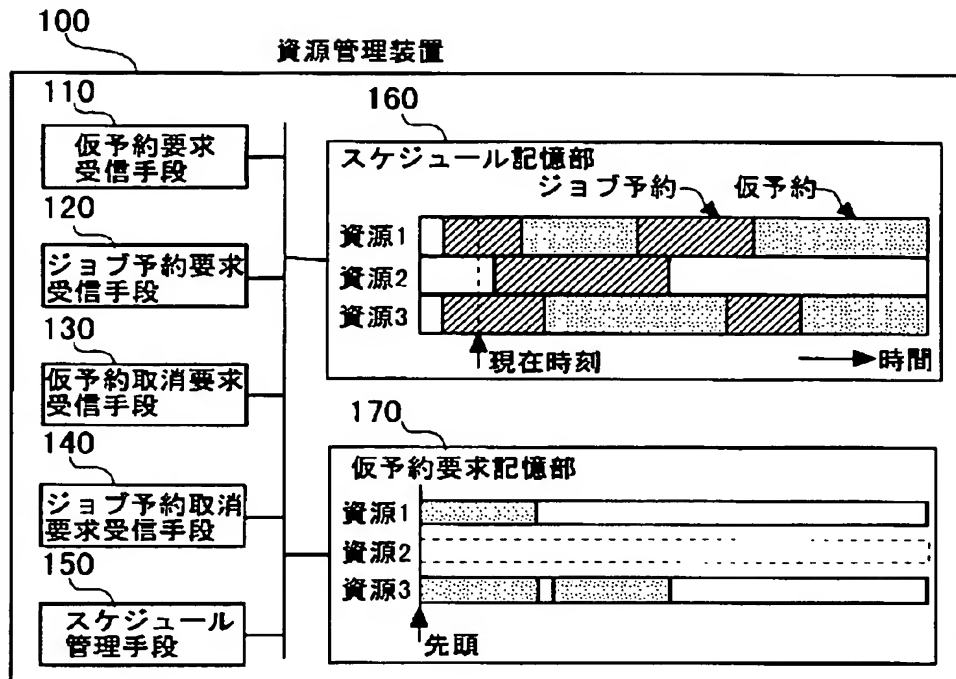


確定したジョブ予約

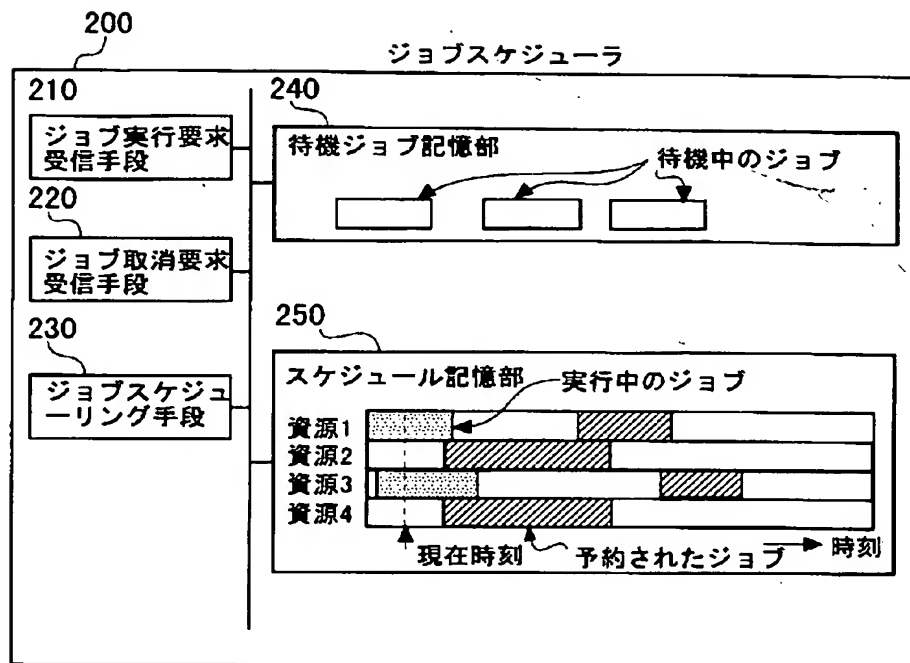
【図 5】



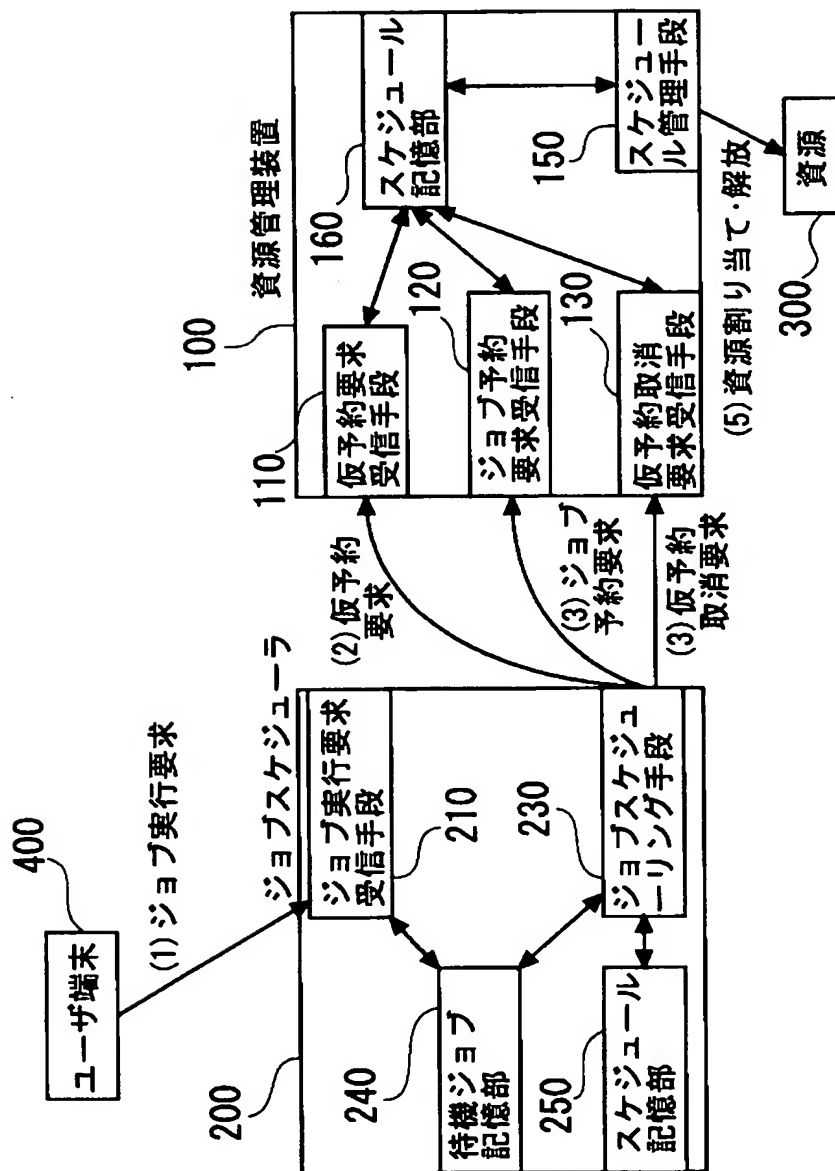
【図 6】



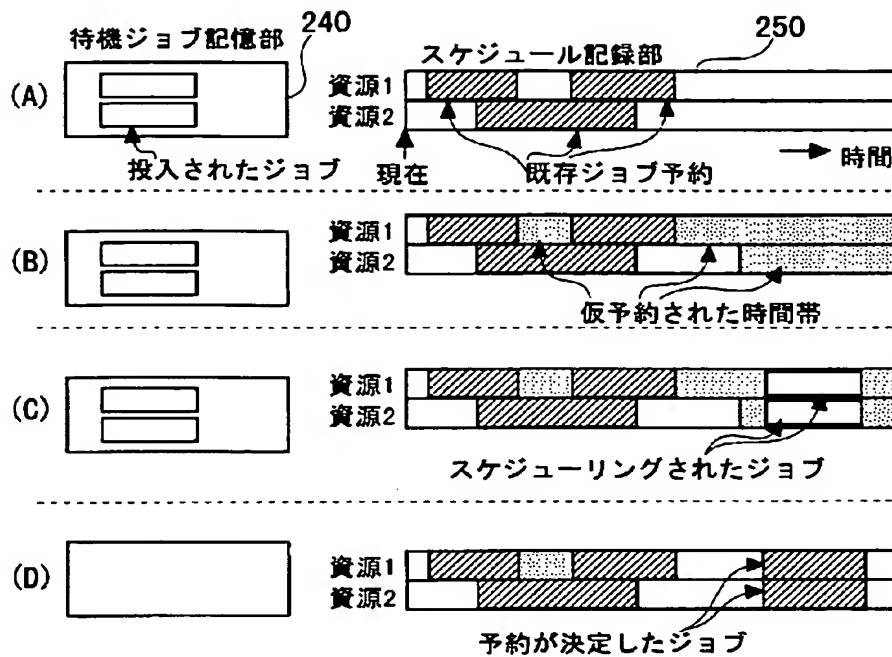
【図 7】



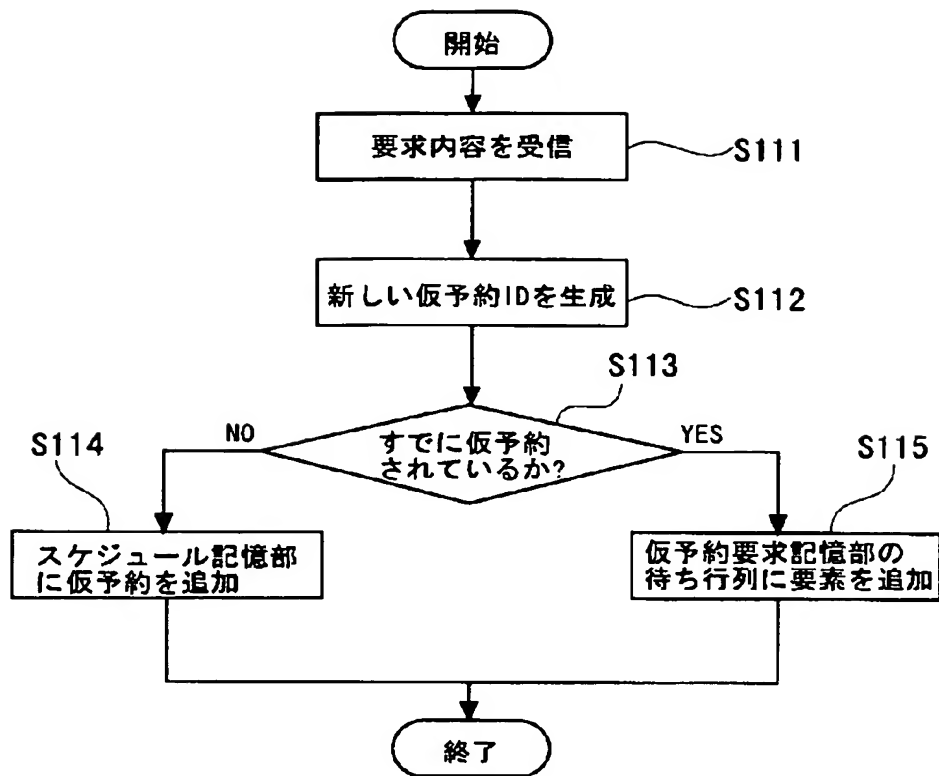
【図 8】



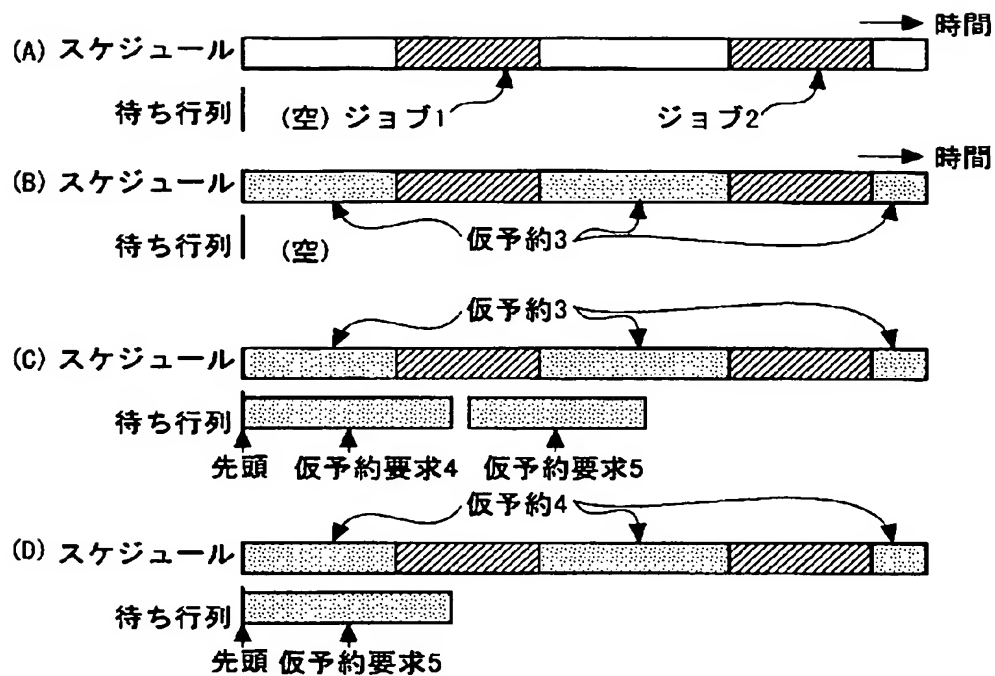
【図 9】



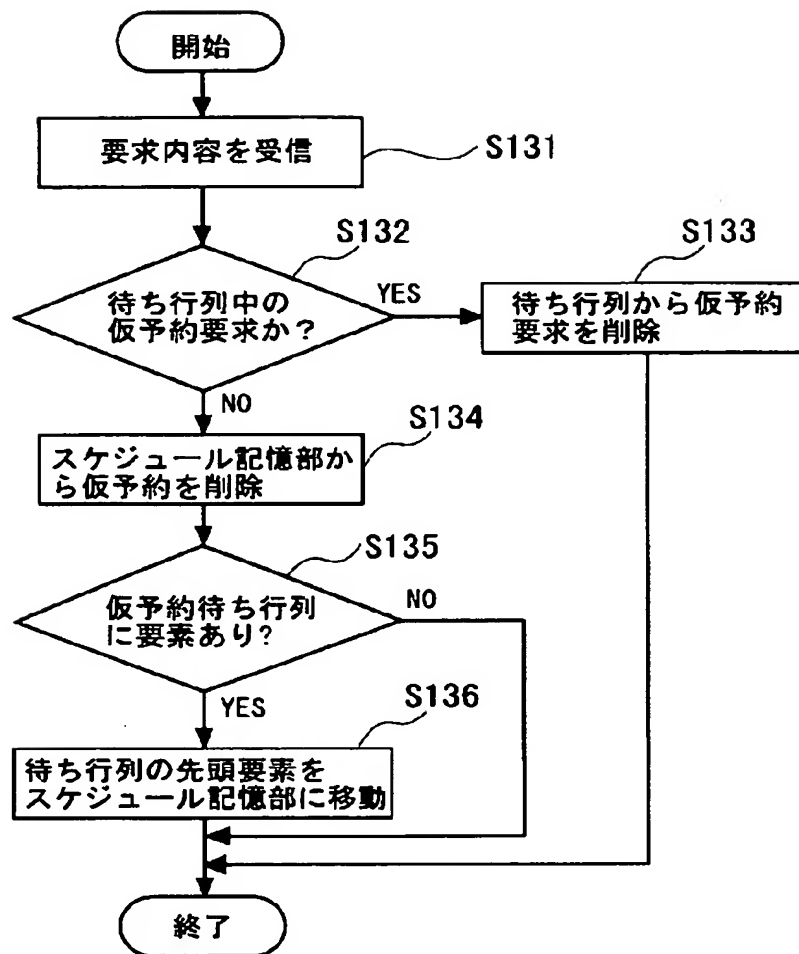
【図10】



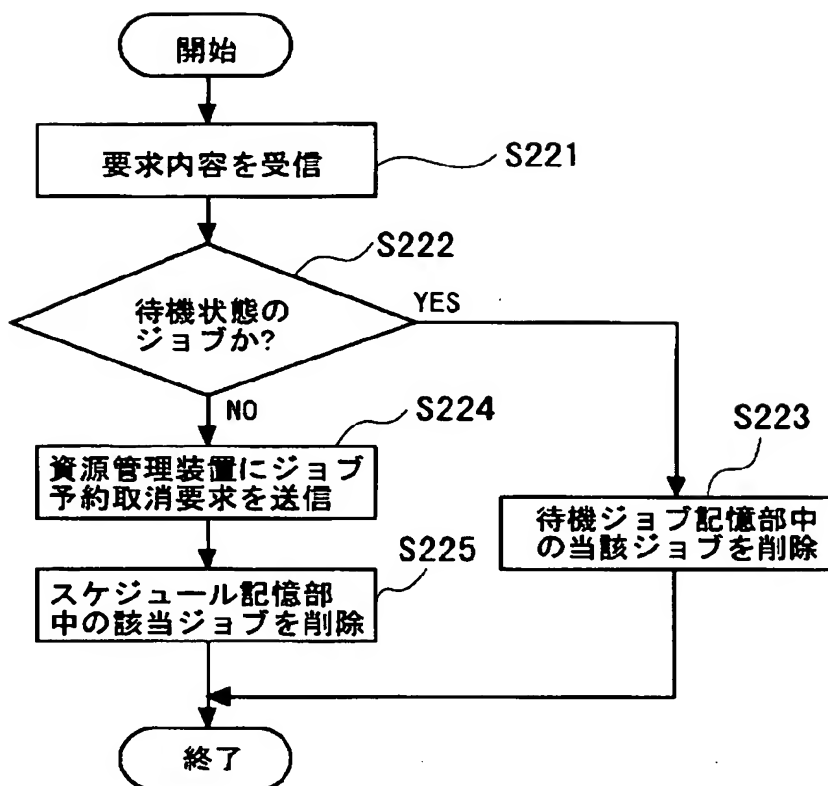
【図 11】



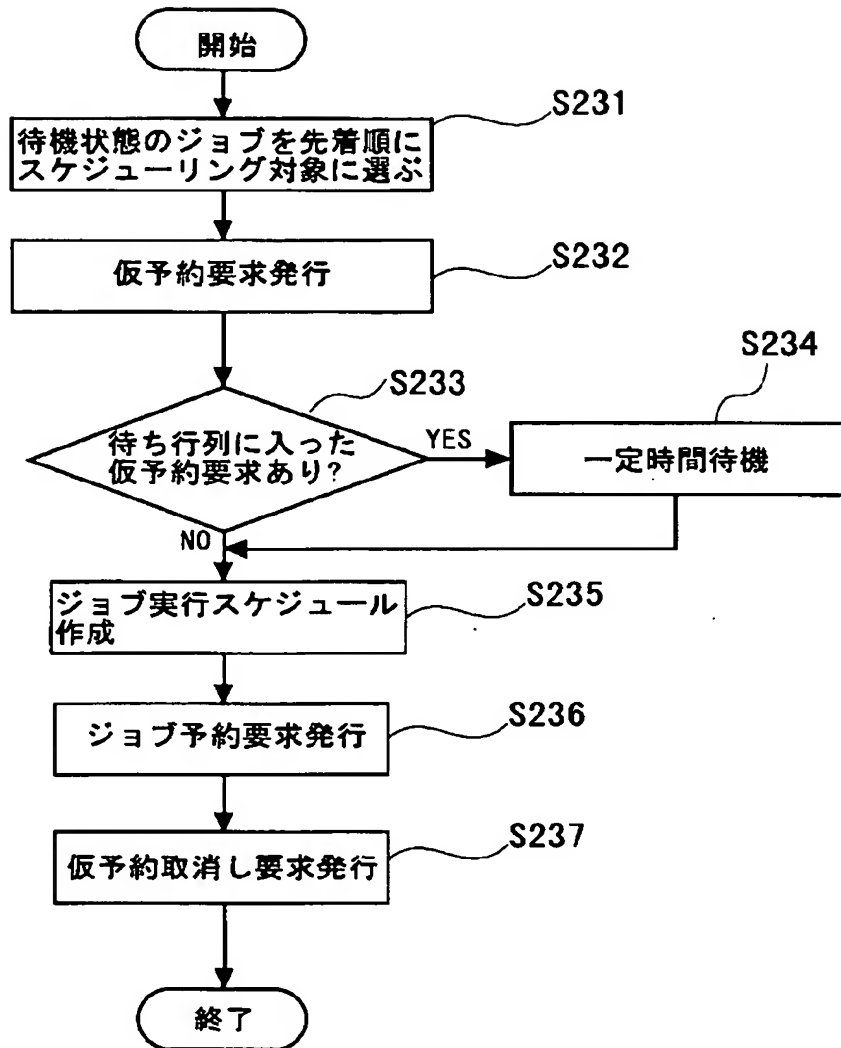
【図 12】



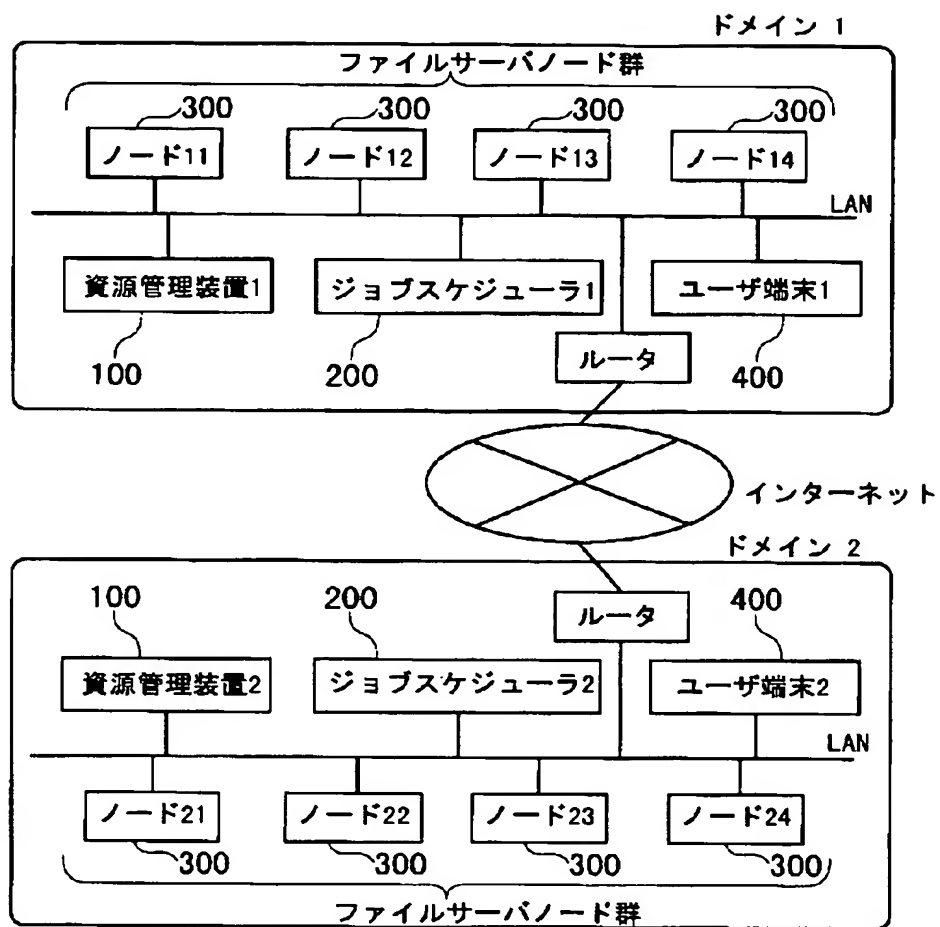
【図 13】



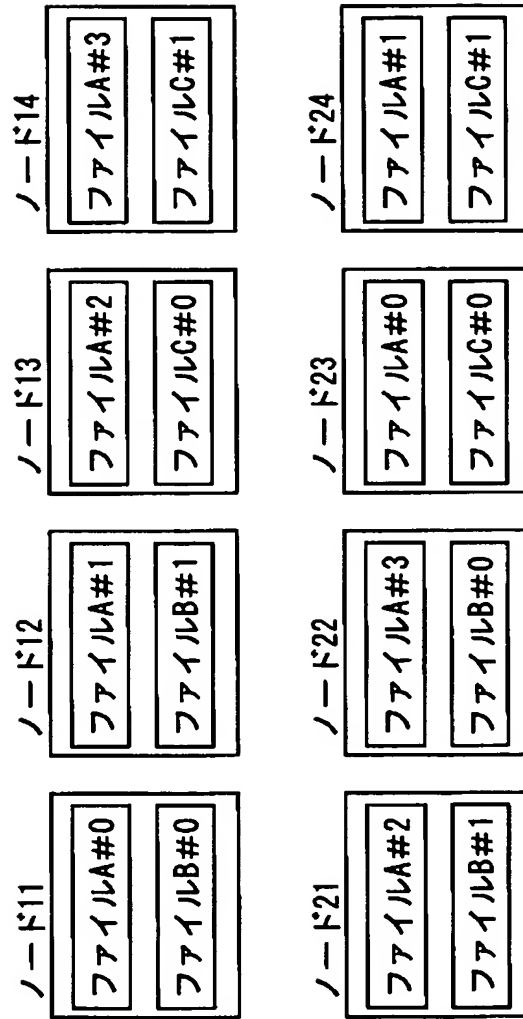
【図14】



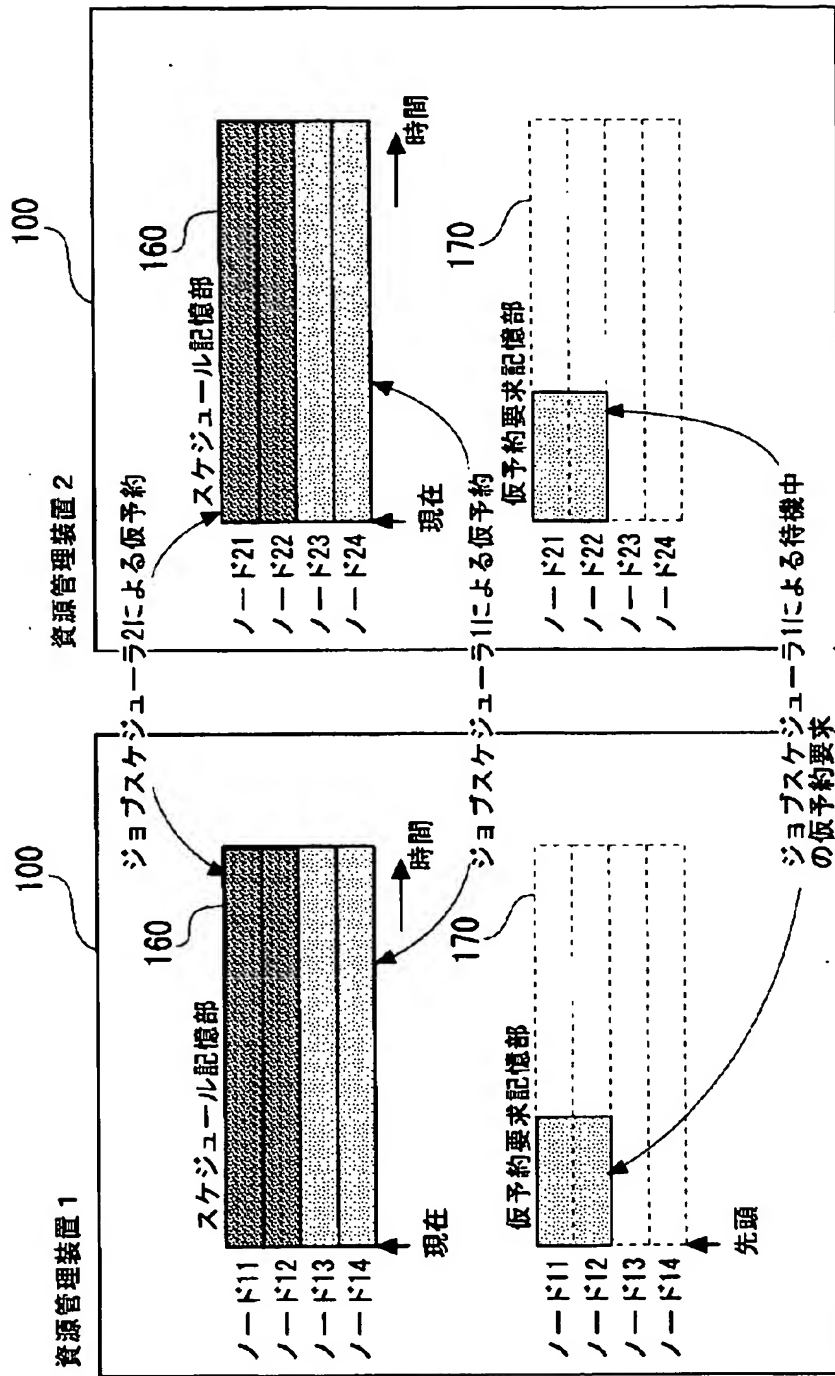
【図 15】



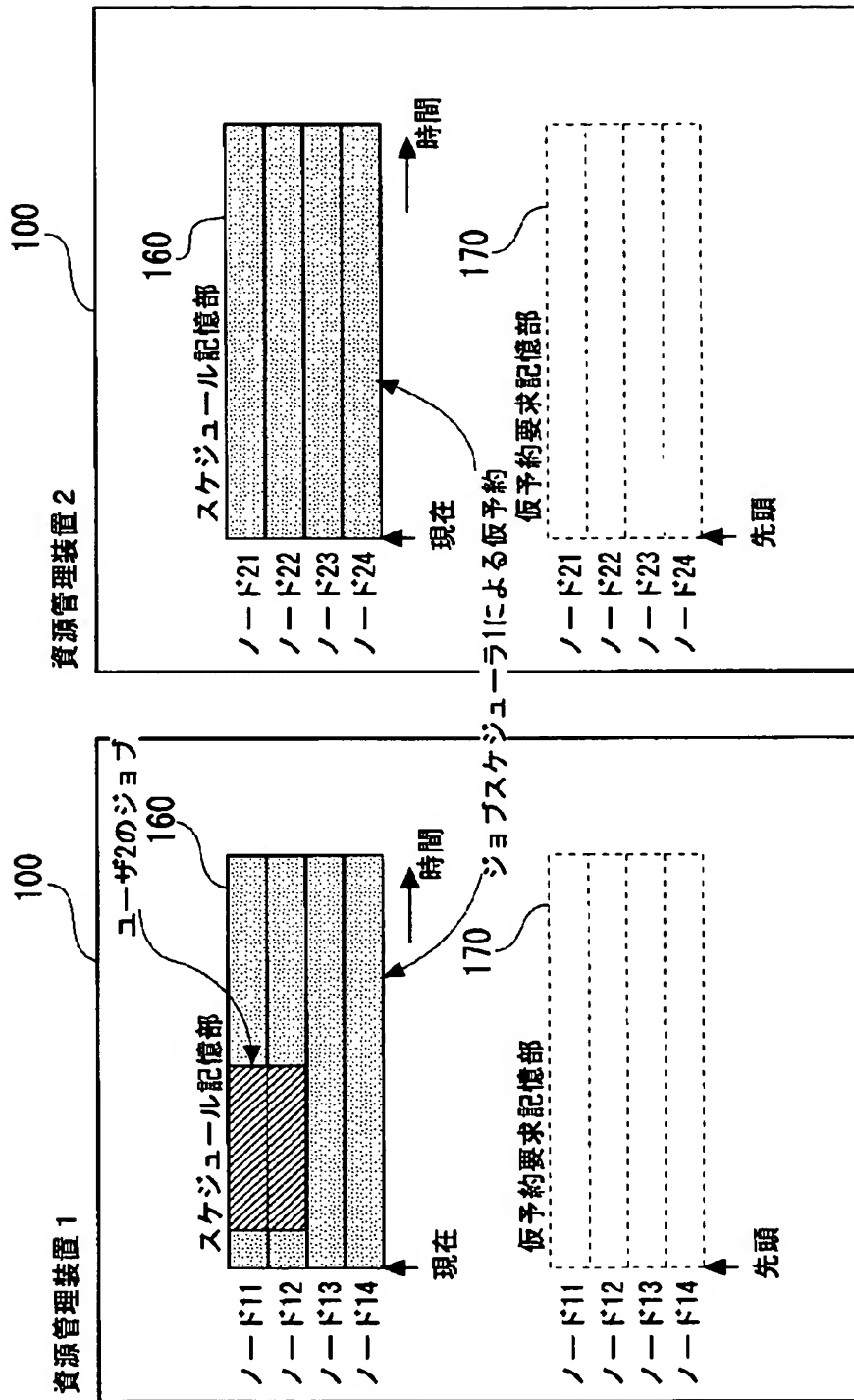
【図 16】



【図 17】



【図 18】



【図 19】

資源管理装置 1 のスケジュール記憶部

ノード11	ジョブ2 処理対象: ファイルB#0	
ノード12	ジョブ2 処理対象: ファイルB#1	
ノード13		
ノード14		

資源管理装置 2 のスケジュール記憶部

ノード21	ジョブ1 処理対象: ファイルA#0	
ノード22	ジョブ1 処理対象: ファイルA#1	
ノード23	ジョブ1 処理対象: ファイルA#2	
ノード24	ジョブ1 処理対象: ファイルA#3	

↑ 現在 → 時間

【図 20】

資源管理装置 1 のスケジュール記憶部

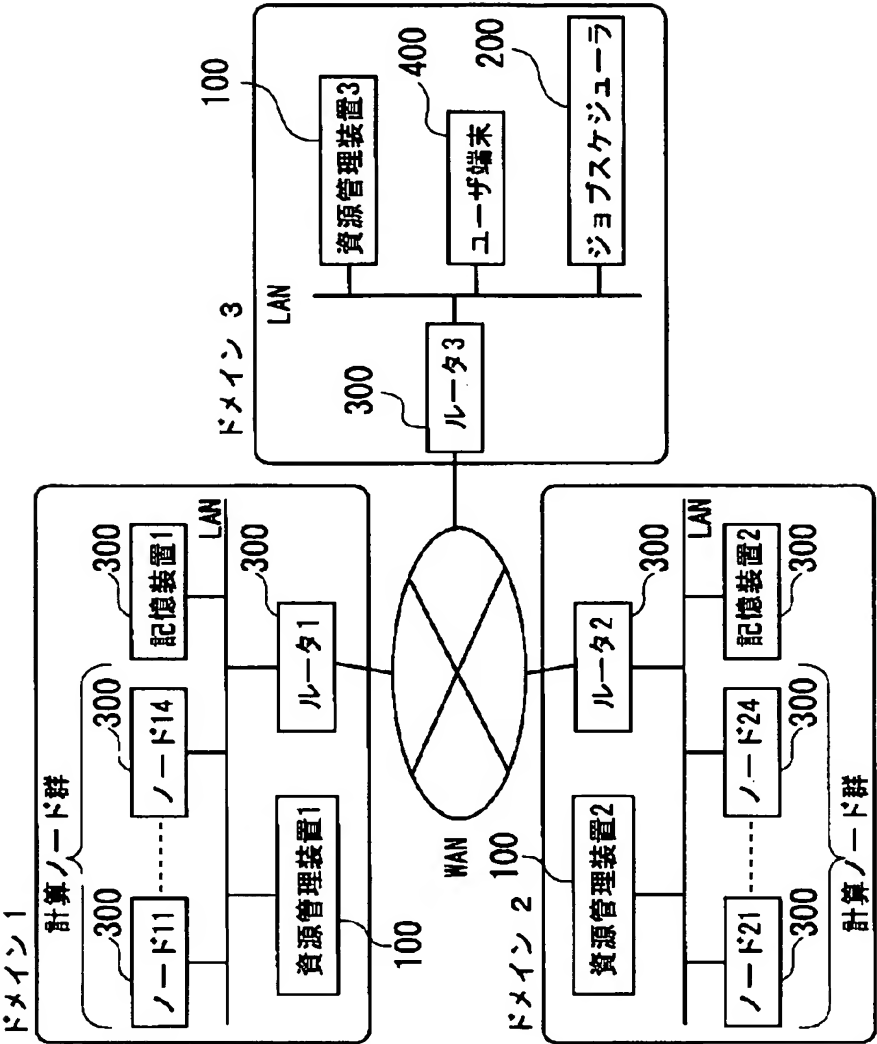
ノード11	ジョブ2 処理対象: ファイルB#0	
ノード12	ジョブ2 処理対象: ファイルB#1	
ノード13	ジョブ1 処理対象: ファイルA#2	
ノード14	ジョブ1 処理対象: ファイルA#3	

資源管理装置 2 のスケジュール記憶部

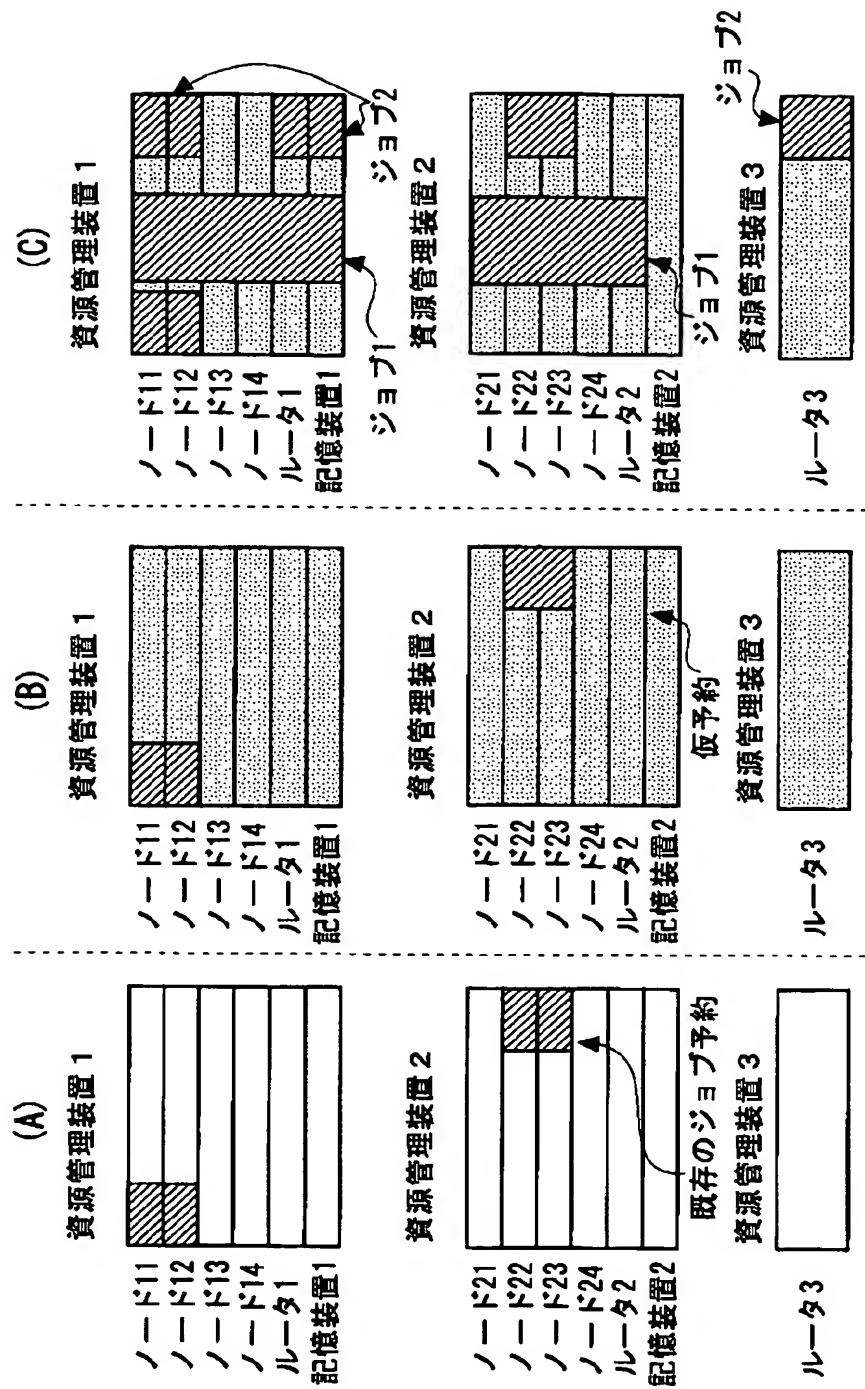
ノード21	ジョブ1 処理対象: ファイルA#0	
ノード22	ジョブ1 処理対象: ファイルA#1	
ノード23		
ノード24		

↑ 現在 → 時間

【図 21】

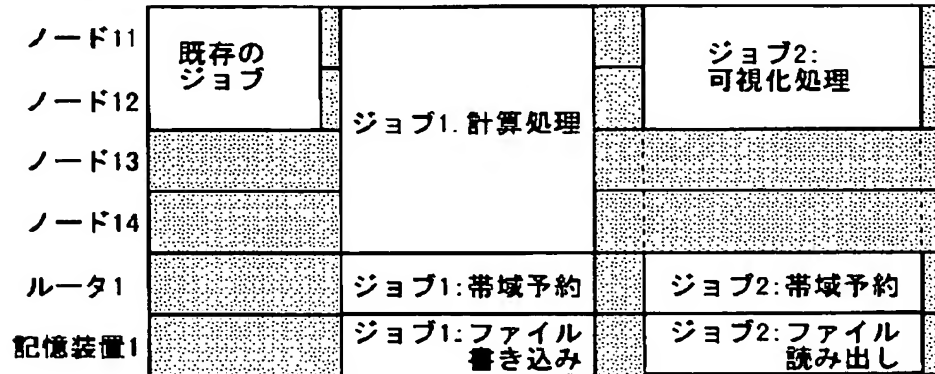


【図 22】

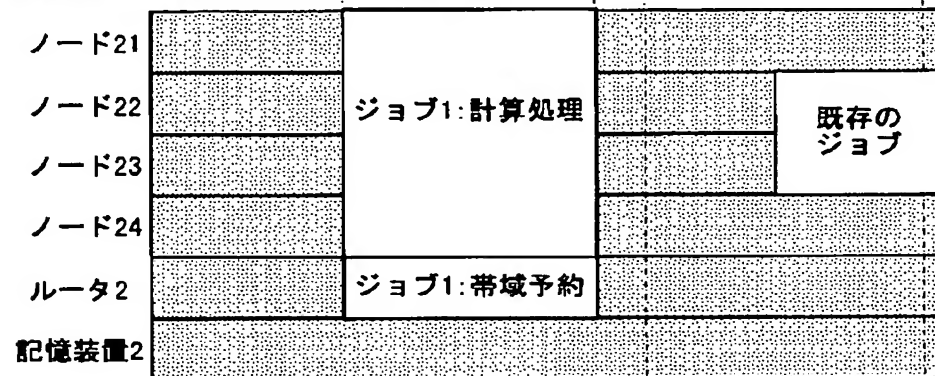


【図 23】

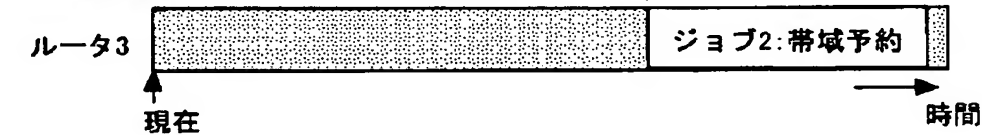
資源管理装置 1 のスケジュール記憶部



資源管理装置 2 のスケジュール記憶部



資源管理装置 3 のスケジュール記憶部



【書類名】 要約書

【要約】

【課題】

複数の資源管理装置によって分散管理される分散システムにおいて、資源を事前に予約するジョブを実行する際に予約の失敗が連続して発生することを避ける装置及び方法の提供。

【解決手段】

資源の利用を予約するために、資源管理装置機構に対して仮予約とジョブ予約の2段階が設けられている。このうち仮予約は短時間のみ有効であり、他の仮予約およびジョブ予約とは共存できない。資源(300)を実際に利用するには、仮予約した後で、ジョブ予約を実行する必要がある。すでに仮予約された資源に、仮予約を試みると、待ち行列に入り、後で仮予約が成功することが保証される。ユーザ端末(400)はジョブスケジューラ(200)にジョブの実行を要求し、ジョブスケジューラは資源管理装置(100)に仮予約操作を実行した後でジョブを割り当てる資源を決定し、ジョブ予約を実行することでジョブの実行に必要な資源を予約する。

【選択図】

図 5

特願 2 0 0 3 - 0 9 3 4 5 9

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 4 2 3 7]

1. 変更年月日

1 9 9 0 年 8 月 2 9 日

[変更理由]

新規登録

住 所

東京都港区芝五丁目 7 番 1 号

氏 名

日本電気株式会社